

Fonds structurels communautaires

“Évaluer les programmes socio-économiques”

Collection
MEANS

Commission européenne





Fonds structurels communautaires

Évaluer les programmes socio-économiques
Conception et conduite d'une évaluation

Collection
MEANS

Volume 1

Commission européenne

Évaluer les programmes socio-économiques

Conception et conduite d'une évaluation

Volume 1

*MEANS est un programme de la Commission européenne.
Il vise à améliorer les Méthodes d'Evaluation des Actions de Nature Structurelle.
Le programme et la Collection MEANS ont été lancés et supervisés
par MM Philippe Goybet et Miguel Angel Benito Alonso
DG XVI/G2, Unité de Coordination de l'Evaluation
Coordination-evaluation@fmb.dg16.cec.be*

*Sa réalisation a été confiée au Centre Européen d'Expertise en Evaluation (C3E)
sous la direction d'Eric Monnier et de Jacques Toulemonde.
Les opinions exprimées dans cet ouvrage n'engagent que les auteurs du texte :
C3E - 13bis Place Jules Ferry - F-69006 Lyon - France - c3e@c3e.fr*

De nombreuses autres informations sur l'Union européenne sont disponibles sur Internet via le serveur Europa (<http://europa.eu.int>), et sur le site Info regio (<http://info regio.cec.eu.int>).

Une fiche bibliographique figure à la fin de l'ouvrage.

Luxembourg:

Office des publications officielles des Communautés européennes, 1999

ISBN 92-828-6627-0

© Communautés européennes, 1999

Reproduction autorisée, sauf à des fins commerciales, moyennant mention de la source.

IMPRIMÉ SUR PAPIER BLANCHI SANS CHLORE

Printed in Italy

Préface



Développer des pratiques appropriées en matière d'évaluation est devenue l'une des actions prioritaires de la Commission afin de contribuer à la bonne utilisation des financements communautaires.

En effet, le respect des principes de bonne gestion financière qu'elle cherche à promouvoir dans son programme SEM2000 (Sound and Efficient Management) repose notamment sur la nécessité de procéder à l'évaluation régulière des actions communautaires.

Les Directions Générales en charge des politiques structurelles possèdent dans

ce domaine une certaine expérience, car elles ont dû intégrer la démarche d'évaluation dès 1988, tout au long du cycle des interventions structurelles communautaires : lors de leur conception sous la forme d'une évaluation prospective, lors de leur mise en œuvre en tant que renforcement du système de suivi et, enfin, à l'issue de celles-ci afin de tirer les enseignements des impacts obtenus.

Cependant, la capacité de l'évaluation à contribuer à l'efficacité et à la transparence de ces actions doit s'appuyer sur la crédibilité des méthodes qu'elle emploie afin qu'elles puissent garantir la validité des résultats.

C'est pourquoi, il y a quatre ans, à l'initiative de la Direction Générale de la Politique Régionale et de la Cohésion (DG XVI), le programme MEANS a été conçu comme une réponse à ce souci de développer des méthodologies et des pratiques d'évaluation conciliant la rigueur scientifique et les spécificités propres aux interventions communautaires, tout en demeurant volontairement accessibles au plus grand nombre.

L'un des résultats les plus tangibles du programme MEANS est la publication de cette collection de six ouvrages rédigés par un groupe d'experts indépendants qui traitent des grands aspects de l'évaluation des programmes socio-économiques dont un bref aperçu est donné au début de ce volume. Cette collection capitalise des expériences, des échanges et des innovations méthodologiques dans le contexte des Fonds Structurels. Elle permettra aux responsables de l'évaluation de développer leurs

pratiques de manière autonome et aux commanditaires de mieux maîtriser cette démarche.

Témoin direct lors des grandes rencontres européennes entre spécialistes et utilisateurs de l'évaluation (Bruxelles 1995, Berlin 1996, Séville 1998) de l'engouement que suscitent ces nouvelles techniques, il m'appartient d'insister sur le fait que le développement de l'évaluation doit aller de pair avec la recherche d'une véritable insertion de ses résultats dans les processus décisionnels afin de savoir si les politiques menées sont équitables, efficaces et adaptées aux besoins.

En répondant à ces exigences, l'évaluation apparaît comme un moyen privilégié offrant tant aux autorités nationales que communautaires, les garanties nécessaires pour que les fonds communautaires soient utilisés au mieux des intérêts des citoyens de l'Union. Les enjeux de la Cohésion Economique et Sociale ainsi que les défis que représente le futur élargissement de l'Union européenne sont à ce prix.

Je me dois enfin de remercier le Comité des Personnalités Indépendantes constitué dès le lancement du programme qui par leur compétence et leur renom ont grandement contribué à asseoir la qualité et la pertinence méthodologique du programme MEANS. Il s'agit de : E. Chelimsky (USA), ex-Présidente de l'«American Evaluation Association» ; K. Kennedy (IRL), Directeur de l'«Economic and Social Research Institute» (ESRI) ; J. R. Cuadrado Roura (E), Président de l'«European Regional Science Association» ; C. Seibel (F), Directeur des Statistiques démographiques et sociales de l'Institut National de la Statistique et des Etudes Economiques (INSEE) ; L. Senn (I), Directeur du Département d'Economie Régionale de l'Université Bocconi ; E. Stern (UK), Directeur de l'unité d'évaluation du Tavistock Institute ; A. Sorber (†) (NL), Chef du «Policy Analysis Department of the Ministry of Finance» ; L. I. Strom (SV), Statens Institut för Regionalforskning (SIR) ; L. Tsoukalis (GR), professeur à l'Université d'Athènes ; H. Wollmann (D), professeur à l'Université de Humboldt à Berlin et membre fondateur de la «European Evaluation Society».

E. Landaburu



Table des matières

La Collection MEANS	5
Présentation du Volume 1	11
I Les enjeux de l'évaluation dans le cadre des Fonds structurels	15
La valeur ajoutée de l'évaluation	17
La spécificité des programmes et de leurs évaluations	27
Le développement des pratiques dans le cadre des Fonds structurels européens	35
II Préparer l'évaluation et définir son contenu	49
Définir le cadre général de l'évaluation projetée	51
Décider le lancement de l'évaluation	67
Passer commande de l'évaluation	79
III Réaliser l'évaluation d'ensemble	89
Analyser la logique et le contenu du programme	91
Evaluer les impacts de l'ensemble du programme	115
Analyser les impacts macroéconomiques	145
Réaliser des analyses plus approfondies	157
IV Utiliser l'évaluation	171
Optimiser l'utilisation de l'évaluation	173
Apprécier la qualité d'une évaluation	181
V Assurer la réussite d'une évaluation	195
Mieux préparer l'évaluation	197
Ouvrir l'évaluation à des partenaires plus nombreux	203
Consolider les techniques d'évaluation	211
Annexes	217
Annexe 1 - Processus d'évaluation ex ante, à mi-parcours et ex post	219
Annexe 2 - Guide pour la préparation d'un cahier des charges	225
Annexe 3 - Références bibliographiques	231



La Collection MEANS

La Commission européenne a investi pendant plusieurs années dans des actions de recherche, de développement, de formation et de communication, en vue d'améliorer la qualité et l'utilité des évaluations réalisées dans le cadre de la Politique de cohésion économique et sociale. Ces travaux ont pris place dans un programme d'ensemble intitulé MEANS, Méthodes d'Evaluation des Actions de Nature Structurale, et initié par la Direction Générale XVI (Politique régionale et cohésion). Ce programme se justifie par l'enjeu que représente l'évaluation dans un domaine qui représente un tiers du budget de l'Union. L'objectif général du programme est de constituer un ensemble cohérent de méthodes pour traiter de façon efficace les problèmes d'évaluation rencontrés dans le domaine des interventions des Fonds structurels.

Les productions du programme MEANS constituent une somme d'expériences, de réflexions et d'innovations méthodologiques qui ont jusqu'à présent fait l'objet de sept cahiers qui préfigurent la présente Collection. En outre, ont été réalisés des bulletins trimestriels, des notes méthodologiques, des séminaires de formation et trois conférences européennes (la dernière, à Séville, a réuni près de 600 praticiens de l'évaluation des programmes des Fonds structurels). Les travaux ont cherché à tirer le meilleur parti des expériences existantes et des innovations récemment apparues, notamment dans les domaines suivants : mesure des effets sur l'emploi, développement des systèmes d'indicateurs, évaluation des effets macroéconomiques. L'évaluation a fait l'objet de nombreuses journées de formation et de sensibilisation au sein de la Commission dans le secteur des Fonds structurels.

Grâce aux évaluations initiées au niveau européen, et à la capitalisation de ces expériences, les acquis sont maintenant suffisamment substantiels pour nourrir une

publication complète et accessible à un large public : une collection d'ouvrages traitant tous les aspects de l'évaluation dans le cadre des politiques structurelles et tirant le meilleur parti possible des textes déjà produits dans le cadre du programme MEANS, remaniés et complétés pour garantir leur cohérence.

Le parti a été pris de concevoir la Collection sans faire trop directement référence aux dispositions réglementaires qui régissent la politique européenne de cohésion pour la période 1994-1999. En effet, l'expérience a montré que le cadre réglementaire change périodiquement alors que les problèmes techniques et les standards professionnels ont une plus grande permanence. Aussi, en vue d'accroître la longévité de la Collection, il a été intentionnellement choisi de la faire porter sur les « programmes de développement socio-économique ». Ce terme, considéré au sens large, s'applique aussi bien aux interventions financées dans le cadre de la politique européenne de cohésion, qu'à toute autre intervention ayant les mêmes caractéristiques essentielles, à savoir : une finalité globale de développement, des objectifs multiples, une dimension partenariale et un contenu multisectoriel. De fait, l'évaluation de tels programmes requiert des méthodes spéciales que l'on ne peut pas trouver de façon simple et exhaustive dans la littérature existante. C'est la justification principale de la Collection MEANS.

Une aide pour ceux qui pilotent ou réalisent les évaluations

La Collection s'adresse à deux types de lecteurs : ceux qui sont concernés par le pilotage des évaluations, d'une part, et par leur réalisation, d'autre part. Les lecteurs de la première catégorie, participent à l'ensemble du processus d'évaluation, y compris aux décisions qui interviennent en amont et en aval de la réalisation du travail. Ce sont principalement les fonctionnaires

européens (rapporteurs des unités géographiques de la Commission), nationaux et locaux chargés de la gestion des programmes financés par l'Union européenne. La Collection doit leur permettre, en particulier lorsqu'il s'agit de fonctionnaires et agents peu formés en ce domaine, de disposer d'un recueil complet de l'état de l'art, en vue d'assurer pleinement leurs missions. Toutefois, les travaux du programme MEANS ont permis des avancées méthodologiques d'ordre plus général qui devraient intéresser les personnes responsables de commandes d'évaluations dans d'autres domaines.

La seconde catégorie de lecteurs à laquelle cette collection est destinée comprend les équipes chargées des travaux d'évaluation proprement dits : centres de recherche publics, laboratoires universitaires, sociétés de conseil privées, consultants indépendants, et unités d'évaluation internes à l'administration.

Par ailleurs, cette série d'ouvrages est susceptible d'intéresser les enseignants et les étudiants dans les programmes académiques traitant d'économie régionale, de politiques économiques, de politiques de l'emploi, de management public, etc.

Cet ouvrage n'a cependant pas une vocation académique. En particulier, la bibliographie est limitée à quelques textes de référence et ne prétend pas à l'exhaustivité.

Tour d'horizon de l'évaluation en six Volumes

La Collection complète comprend six Volumes conçus pour être complémentaires et fournir les éléments indispensables pour qu'un professionnel puisse perfectionner sa pratique de façon autonome. En outre, chaque ouvrage a été construit de telle manière qu'il puisse être lu et compris indépendamment des autres Volumes de la Collection. Outre le présent ouvrage, la Collection est organisée selon la structure suivante :

Le deuxième Volume traite du choix et de l'utilisation des indicateurs pour le suivi et l'évaluation. Il montre

comment les indicateurs peuvent être utilisés pour appuyer des décisions politiques, pour renforcer la gestion et pour mesurer la performance globale des programmes. L'ouvrage propose un cadre méthodologique, de nombreux exemples d'indicateurs, et un guide pour l'appréciation de la qualité des indicateurs. Il vise à faire progresser les pratiques en termes de quantification des indicateurs et de comparaisons entre programmes.

Le troisième Volume présente les **principales techniques et outils d'évaluation** adaptés aux programmes socio-économiques. Ces outils sont issus de six ou sept disciplines scientifiques différentes: économie, sociologie, science politique, management, géographie, etc. C'est en quelque sorte une encyclopédie technique présentée sous forme de fiches dont chacune comprend la description générale d'un outil, les principes de sa mise en œuvre, une liste de ses points forts et faibles, un exemple d'application et une bibliographie des ouvrages de référence. En plus de donner un aperçu des possibilités d'utilisation de chacun des outils pris de façon isolée, le document inclut un guide pour le choix des outils et pour leur combinaison en fonction des questions posées et du contexte. Un tel ouvrage à caractère multidisciplinaire n'a pas d'équivalent dans la littérature existante et sera sûrement un des apports les plus originaux de cette Collection.

Plusieurs **solutions techniques pour évaluer dans un cadre de partenariat**, font l'objet d'un quatrième Volume. Cet ouvrage est rendu nécessaire par le fait que les programmes de développement socio-économique, constituent des objets d'évaluation inhabituels. Premièrement, ils impliquent toujours plusieurs autorités publiques dans le cadre d'un accord partenarial qui n'empêche pas l'existence de points de vue légitimement différents. Leurs objectifs sont généralement nombreux et pas toujours suffisamment clairs et cohérents pour fonder l'évaluation. Deuxièmement, les programmes

interviennent dans de multiples secteurs, ce qui multiplie le travail d'observation et d'analyse dans des proportions considérables, et qui pose des problèmes au moment de rédiger une synthèse de façon transversale aux multiples composantes du programme. Enfin, chaque autorité publique devant rendre des comptes à « ses administrés », chaque niveau de collectivité introduit dans l'évaluation ses propres critères et priorités qu'il faut prendre en compte. On ne trouve pas dans les manuels d'évaluation existants une solution complète à ces problèmes. Grâce à plusieurs évaluations pilotes, le programme MEANS a testé et mis au point quatre techniques qui peuvent être réunies pour former une solution méthodologique complète. Le Volume décrit le mode d'emploi de ces quatre techniques : la cartographie conceptuelle des impacts, la matrice des impacts croisés, l'échelle de notation et l'analyse multicritère.

Le cinquième Volume traite de l'évaluation transversale des impacts sur l'environnement, l'emploi et autres priorités d'intervention. En effet, l'évaluation des programmes financés par les Fonds structurels ne peut pas se limiter à l'examen de leur seul impact global. Elle se doit d'analyser également la prise en compte d'une série de priorités européennes aussi diverses que l'emploi, le maintien de l'emploi, l'égalité des chances entre les femmes et les hommes, l'environnement, la compétitivité des PME, etc. Ces priorités, ou thèmes prioritaires, sont souvent rappelées dans les documents opérationnels sans toutefois faire l'objet d'une intégration véritable (*mainstreaming*). Il est admis que l'évaluation de tels impacts transversaux, appelée évaluation thématique, est difficile à réaliser et que les difficultés sont accrues lorsque ces impacts sont indirects. Cet ouvrage fournit un cadre méthodologique pour aider à identifier les impacts potentiels d'un programme, à les prendre en compte y compris pour les mesures qui n'y font pas référence, et à identifier et définir des indicateurs appropriés pour suivre et analyser empiriquement leur réalité auprès des publics concernés. Il offre les éléments nécessaires à

l'analyse et au choix de meilleures solutions méthodologiques adaptées aux évaluations thématiques.

Enfin, le dernier Volume vient clore la Collection avec un **Glossaire des concepts et termes techniques**. La rédaction de la Collection a imposé un travail très rigoureux d'homogénéisation et de définition des termes employés et la constitution d'un index spécifique à chaque Volume. Le glossaire récapitule plusieurs centaines de termes tirés des différents index et propose des définitions, traductions et illustrations. Il répond à un réel besoin en ce sens que la bonne qualité d'une évaluation dépend en partie de la qualité de la communication entre les partenaires impliqués dans la commande, d'une part, et entre le commanditaire et l'équipe d'évaluation, d'autre part. Dans le cas de la politique européenne de cohésion économique et sociale, l'interaction est rendue difficile parce que l'évaluation est conduite dans un contexte international et intersectoriel impliquant des personnes qui ne partagent ni la même langue maternelle, ni les mêmes références professionnelles. Par ailleurs, le développement récent de l'évaluation fait qu'un grand nombre d'acteurs n'ont pas la connaissance du vocabulaire standard de l'évaluation, lequel n'est d'ailleurs pas toujours stabilisé. Il en résulte des difficultés de communication qui peuvent compromettre gravement la qualité et l'utilité des travaux d'évaluation. Ce Volume couvre, de façon générale, tous les thèmes abordés dans le reste de la Collection. Dans les situations fréquentes où plusieurs définitions coexistent à la Commission et/ou dans la communauté scientifique, il fait un choix unique et clair tout en mentionnant les principales définitions concurrentes en vigueur dans d'autres contextes sectoriels ou nationaux.

Présentation du Volume 1

Ce premier Volume de la Collection MEANS sert à « entrer » à la fois dans la Collection et dans le domaine de l'évaluation des programmes socio-économiques. En plus des questions traditionnellement traitées dans les guides d'évaluation actuellement disponibles (*General Accounting Office*, 1991 ; Commission Européenne, 1997 ; Conseil Scientifique de l'Evaluation, 1994 ; bibliographie en annexe 3), il présente et traite de façon opérationnelle les questions spécifiques aux programmes socio-économiques telles que l'interaction entre les différents niveaux de gestion (projet, programme, politique), le partenariat entre plusieurs niveaux de gouvernement ou encore, l'évaluation d'ensemble des programmes complexes et les analyses plus approfondies d'actions ou de thèmes spécifiques.

Cet ouvrage doit permettre au lecteur de découvrir et de comprendre les enjeux et la valeur ajoutée de l'évaluation tout en lui indiquant ce qu'il peut ou ne peut pas attendre d'une évaluation à un moment et dans un contexte donnés. Il apporte les éléments nécessaires au pilotage d'une évaluation de A à Z : répartir les responsabilités de l'évaluation, définir sa finalité et les grandes lignes de sa méthode, fixer son calendrier et son budget, apprécier la qualité des travaux et enfin gérer l'utilisation de ses conclusions. Il est applicable de façon générale aux évaluations qui interviennent avant, pendant ou après un programme, autrement dit aussi bien pour l'analyse prospective des effets attendus, des besoins et de la stratégie que pour l'analyse rétrospective des impacts observables sur le terrain. Les modèles d'organisation proposés ont une portée générale et leur ambition ne se limite pas aux schémas organisationnels actuels qui sont amenés à évoluer dans le cadre de la nouvelle génération de programmes après 1999.

On notera que ce premier Volume a un rôle d'introduction générale à l'évaluation des programmes de développement socio-économiques. Il fait la synthèse de l'ensemble de la série et il est destiné à un public moins spécialisé que les Volumes suivants, plus techniques et destinés à un public plus averti.

Ce Volume a été construit de façon à ce que le lecteur puisse en appréhender le contenu en fonction de ses intérêts propres. Les encadrés ci-dessous proposent plusieurs modes de lecture permettant d'aller rapidement à l'essentiel.

Encadré 1 - Pour ceux qui veulent mieux connaître l'évaluation en allant rapidement à l'essentiel

Pour une vue d'ensemble rapide :

- *Parcourir les messages qui sont placés en italique en tête des chapitres et qui résument brièvement les points clés de l'ouvrage – commencer page 17*

Pour le lecteur qui connaît peu les interventions structurelles européennes :

- *Lire attentivement la première Partie « Les enjeux de l'évaluation dans le cadre des Fonds structurels »*

Pour le lecteur qui doit passer commande d'une évaluation :

- *Passer directement à la Partie II « Préparer l'évaluation et définir son contenu »*

Pour ceux qui conçoivent, réalisent ou surveillent des travaux de collecte et d'analyse de données :

- *Lire plus particulièrement la Partie III « Réaliser l'évaluation d'ensemble »*

Pour le lecteur qui désire tirer le meilleur parti possible d'une évaluation :

- *Des conseils et recommandations sont présentés aux Parties II et IV*

Pour réfléchir au perfectionnement des pratiques d'évaluation :

- *Voir les suggestions et recommandations faites en conclusions du Volume à la Partie V*

Encadré 2 - Pour ceux qui doivent concevoir et réaliser une évaluation

Évaluation ex ante :

- Consulter l'annexe 1 - Processus d'évaluation - colonne « ex ante »
- Suivre les symboles ◀ placés en début de section ou de paragraphes

Évaluation à mi-parcours :

- Consulter l'annexe 1 - Processus d'évaluation - colonne « mi-parcours »
- Suivre les symboles ▶ placés en début de section ou de paragraphes

Évaluation ex post :

- Consulter l'annexe 1 - Processus d'évaluation - colonne « ex post »
- Suivre les symboles ▶ placés en début de section ou de paragraphes

I

Les enjeux de l'évaluation dans le cadre des Fonds structurels



La valeur ajoutée de l'évaluation

Evaluer une intervention publique, c'est juger de sa valeur au regard de critères explicites et sur la base d'informations spécialement rassemblées et analysées. Au-delà de cette définition générale et forcément abstraite, la pratique de l'évaluation s'est développée dans les pays de l'Union européenne selon plusieurs conceptions : l'évaluation managériale qui vise à améliorer la gestion, l'évaluation démocratique utilisée pour rendre des comptes aux citoyens et l'évaluation pluraliste qui cherche à trouver des accords entre les partenaires de l'action publique. On retrouve ces trois conceptions dans l'expérience d'évaluation des politiques de l'Union européenne, avec une prédominance de l'une ou de l'autre selon qu'il s'agit d'une évaluation *ex ante*, *ex post* ou à mi-parcours.

■ Une démarche récente, propre au secteur public

On évalue pour vérifier la raison d'être d'une intervention publique, pour identifier les réussites reproductibles ou les échecs à ne pas renouveler, pour rendre des comptes aux citoyens. Les organisations du secteur privé n'ont pas véritablement à se soucier de justifier leurs actions ou d'en rendre compte car elles sont jugées par le marché. Evaluer est par contre une nécessité pour les organisations du secteur public.

● Origines de l'évaluation

L'évaluation est en relation avec l'analyse des actions publiques et plus généralement avec les sciences sociales. Elle utilise de façon systématique les méthodes de l'économie et de la sociologie, mais aussi de la psychologie sociale, de la géographie économique, de l'anthropologie, etc. Elle s'est développée, en tant qu'instrument de connaissance, en même temps qu'est apparue la collecte régulière des données socio-démographiques et économiques (état civil, recensement, etc.). Par contre, l'évaluation des programmes, en tant qu'aide au processus décisionnel, est plus récente. Elle apparaît dans les années 1950, alors qu'intervient aux USA une remise en cause de l'efficacité de l'administration et de sa neutralité. Elle fait alors partie des efforts déployés en vue de rationaliser la planification et l'affectation des ressources. C'est la naissance du système de *PPBS* (*Planning, Programming, Budgeting System*) et sa diffusion dans la plupart des pays européens (Voir Rist, 1990, pour une analyse de l'évaluation dans sa perspective historique).

Des évaluations commencent à être menées sur une vaste échelle aux USA à partir des années 60. Elles portent principalement sur les innovations dans les domaines de l'éducation, des services sociaux, de l'hygiène publique et de la prévention du crime. De nombreux programmes sociaux deviennent pour la communauté scientifique américaine de véritables laboratoires expérimentaux dont on attend un accroissement des connaissances et une amélioration de l'action des pouvoirs publics. Mais dès les années 70, les responsables administratifs prennent conscience du caractère contraignant et factice de l'approche de type "expérimentation sociale » et de la norme du PPBS.

L'évaluation a ensuite été davantage perçue comme un moyen de gérer les nécessaires restrictions budgétaires. Ainsi en est-il en Grande-Bretagne, où le développement de l'évaluation est lié au souci de limiter les dépenses publiques qui a accompagné la réforme du service public dans les années 80.

L'évaluation fait désormais partie d'une nouvelle conception de la gestion publique, plus décentralisée et plus responsable. Elle est le reflet d'un certain nombre de préoccupations qui peuvent être déclinées selon trois axes principaux. En premier lieu, on trouve le besoin d'estimer le plus objectivement possible les effets des politiques ou des programmes sur la société et de comprendre les logiques de leur fonctionnement (visée cognitive). En deuxième lieu, il y a le souci d'aider les commanditaires politiques et toutes les parties prenantes à porter un jugement sur la valeur desdits programmes (visée normative). Enfin, en troisième lieu, il y a la nécessité de contribuer pratiquement à l'amélioration des programmes en éclairant leurs différents protagonistes sur le sens, les conditions et les conséquences de leurs actions et décisions (visée instrumentale).

L'évaluation rassemble les informations existantes en un tout cohérent et, le cas échéant, elle en produit de nouvelles afin d'accroître la connaissance des responsables sur les réalités sociales et économiques dans lesquelles s'inscrivent les interventions publiques. En outre, l'évaluation est un instrument au service des décisions stratégiques et opérationnelles : son utilité pratique peut se définir comme une mise en relation des objectifs, des moyens, des résultats et des effets des interventions en vue d'améliorer l'efficacité du programme et son efficacité. Enfin, l'évaluation introduit la notion de valeur, et *a fortiori* de jugement de valeur, dans les pratiques administratives. Cette dimension que l'on peut qualifier de « politique » fait toute l'originalité de l'évaluation et suppose l'émergence d'une nouvelle profession à part entière, distincte par exemple de celles de la recherche et de l'étude.

● Nécessité pour le secteur public

Pour mieux comprendre la nécessité de l'évaluation dans le secteur public, et notamment dans la phase ex post lorsqu'elle cherche à mesurer les impacts d'une activité publique, on peut faire un parallèle avec le rôle du marché pour une entreprise privée qui propose des produits et des services à ses clients. En décidant d'acheter ces produits ou de s'adresser à un concurrent, chaque client intègre tous les avantages, les coûts et les inconvénients que son achat va lui apporter à court et à long terme. Les comportements d'achat des clients sont autant de signaux qui dictent à l'entreprise ce qu'elle doit et ne doit pas faire. Finalement, l'entreprise n'a pas besoin de justifier son existence. Elle survit si elle est financièrement viable, c'est à dire si elle a des effets utiles pour ses clients.

Par analogie avec l'entreprise privée, une organisation publique produit également des équipements ou des services. Elle propose ses services (ou quelquefois les impose), non pas à des clients, mais à des destinataires, à des usagers, à des assujettis. Si les personnes qui participent à une action publique en sont satisfaites, cela ne suffit généralement pas pour dire que l'action est réussie. Au contraire, les gestionnaires publics cherchent souvent à produire des effets lointains et globaux, par exemple le développement socio-économique durable, qui ne sont pas synonymes de satisfaction des destinataires. L'évaluation permet de connaître et de juger l'ensemble des effets d'une action publique précise, effets directs et indirects, effets positifs et négatifs. La sanction électorale est certes un signal important pour tout gouvernement, mais son inconvénient majeur est le caractère non discriminant du jugement qui est ainsi exprimé. Seule la démocratie directe telle qu'elle est pratiquée par les citoyens helvétiques permet, dans une certaine mesure, l'expression d'un jugement intervention par intervention (l'évaluation n'en est pas moins développée en Suisse !). Les démocraties représentées au sein de l'Union européenne sont plutôt de type indirecte ; l'évaluation joue donc un rôle crucial, à la fois comme signal pour montrer ce qu'il convient de faire et de ne pas faire, et comme instrument capable de justifier l'existence de l'action publique aux yeux des citoyens.

■ Les développements de l'évaluation en Europe

● Différentes conceptions dans les pays européens

Le développement européen de l'évaluation a commencé dans les pays du Nord (Royaume-Uni Suède, Danemark, Pays-Bas) où elle est maintenant

bien diffusée au sein des administrations. Tous les pays européens sont maintenant concernés, avec des adaptations nationales liées à des attitudes culturelles différentes à l'égard du rôle du gouvernement et un rapport différent entre les pouvoirs législatif et exécutif.

Les orientations prises au sein des pays européens peuvent être classées dans trois grandes catégories. Aux deux orientations traditionnelles que l'on peut qualifier de « managériale » et « démocratique », s'est ajoutée une troisième que l'on nommera « pluraliste ». Ces trois orientations, ou encore conceptions, sont différentes sans être exclusives les unes des autres. L'encadré ci-dessous les décrit succinctement en montrant l'ensemble des principes et des préconisations qui les caractérisent respectivement (Duran, 1992).

Encadré 3 - Trois conceptions de l'évaluation			
	Managériale	Démocratique	Pluraliste
<i>Projet global</i>	Optimisation des ressources publiques	Impliquer les citoyens et leur rendre des comptes	Résolution collective des problèmes
<i>Déontologie</i>	« Évaluateurs » indépendants (experts)	Jugement par des représentants élus	Jugement par les protagonistes de l'action
<i>Standards</i>	Objectivité	Publicité de la démarche	Impartialité

La problématique générale qui sous-tend l'évaluation managériale est de faire face à la limitation des recettes publiques. La question sous-jacente peut se formuler ainsi : comment justifier les arbitrages entre les différentes politiques sectorielles ? La démarche ici privilégiée est celle, interne à l'administration, d'« optimisation » des moyens budgétaires existants.

La dimension politique de l'acte d'évaluation est passée sous silence. « L'évaluateur », selon la terminologie officielle, est avant tout un méthodologue dont la scientificité fonde la légitimité. La déontologie est basée sur deux principes : l'indépendance de l'évaluateur et l'objectivité de la méthode. La référence à la rigueur scientifique constitue en quelque sorte le gage de l'indépendance de l'évaluation à l'égard du politique.

Exemple : Le « Value for Money » britannique

Au cours des années 80, de multiples fonctions de l'administration britannique ont été privatisées ou déléguées à des agences semi-indépendantes. Parallèlement, tant au niveau de l'administration centrale qu'à celui des citoyens, les exigences d'efficacité de ces services se faisaient croissantes. Cette tendance apparaissait encore plus évidente dans le cadre du développement, au cours de cette période, d'une culture managériale au sein de l'administration fondée sur des évaluations de type « Value For Money » et la diffusion d'orientations techniques sur le mode d'évaluation de ce principe dans le « Green Book » du Ministère des Finances.

La deuxième conception « démocratique », manifeste la prise de conscience de la crise des systèmes traditionnels de représentation (désaffection des partis politiques, syndicats, etc.). Dans cette conception, les pouvoirs publics se doivent d'être pleinement responsables (*accountable*) de leur action vis-à-vis des citoyens. L'accent est donc mis sur la nécessité pour l'évaluation de renforcer les contre-pouvoirs vis-à-vis de l'administration et plus généralement de l'exécutif, accusé de monopoliser l'expertise.

L'évaluation, définie comme un « jugement porté sur la valeur des politiques publiques » doit nourrir le débat démocratique. Il est recommandé que ces évaluations soient dirigées par des instances d'évaluation comprenant des personnalités démocratiquement élues. Les professionnels de l'évaluation sont au service de tels comités sur les aspects techniques uniquement. Le processus doit être rendu public afin de s'assurer de son caractère démocratique. D'une manière générale, on observe que ce modèle tend à se développer, en particulier dans le cadre de la modernisation des pratiques législatives ou politiques qui privilégie aujourd'hui la notion de « lois d'orientation » et de *sunset legislation*, autrement dit de lois à visée prospective et qui doivent faire l'objet d'un vote de confirmation au vu des résultats obtenus à l'issue des premières années de mise en œuvre.

Le dernier type d'évaluation est désigné sous le nom d'évaluation pluraliste ou d'évaluation participative, ou encore dans la littérature américaine de *stakeholder evaluation*. Cette approche essaie de trouver un compromis entre les nécessités techniques qui justifient la première conception et les exigences politiques qui inspirent la deuxième.

Exemple : Les conférences de consensus au Danemark

Parmi les expériences démocratiques, une pratique danoise fait référence. Les « Conférences Publiques de Consensus » font partie du paysage politique depuis plus de dix ans. Le Danish Board of Technology fait appel à des citoyens ordinaires afin de s'exprimer et débattre, avec des experts et représentants des divers intérêts en cause, sur des sujets technologiques particulièrement controversés (par exemple, les manipulations génétiques sur des animaux, les tests susceptibles d'être pratiqués lors d'embauche au sein des entreprises ou de la signature de contrats d'assurance.) Ces conférences, étalées sur trois week-ends, permettent aux citoyens de s'exprimer et dialoguer avec des experts. Les rapports qui sont ainsi produits visent à éclairer l'opinion publique et faciliter le débat démocratique. Sans se substituer à la décision politique, elles jettent les bases du débat entre exécutif et législatif et s'intègrent dans le processus de décision.

Dans cette conception, l'évaluation fait partie d'un processus collectif de résolution de problèmes. Sur la base d'informations fiables et acceptées par tous, les jugements de valeur sont formulés en recherchant un consensus au sein d'une instance pluraliste d'évaluation qui comprend des responsables politiques et administratifs, ainsi que des porte-parole des publics concernés.

Exemple : Des instances pluralistes pour évaluer

En France, des instances d'évaluation ad hoc sont fréquemment constituées pour diriger les travaux d'évaluation. Ainsi, dans le cadre de l'évaluation nationale de la prévention des risques naturels, l'instance d'évaluation comprenait des acteurs d'origines diverses : ministères, autorités locales, agences publiques, associations... Dans le cadre de politiques cofinancées par l'Etat et les régions, le pilotage de certaines évaluations est assuré par des responsables politiques et administratifs des deux niveaux, en association avec des représentants des publics concernés.

On peut faire un rapprochement entre les différentes conceptions énoncées ci-dessus et les « quatre générations » proposées par les ouvrages publiés aux Etats-Unis qui font référence aux différentes phases de développement des pratiques dans ce pays depuis les années 50. Les évaluations dites de « première génération » se sont concentrées sur la description des effets produits dans la société par les interventions publiques. La seconde génération s'est ensuite intéressée à l'analyse des relations de causalité et à l'attribution des effets aux interventions. La troisième génération a commencé à vérifier systématiquement l'atteinte des objectifs visés, ce qui la rend peu différente de l'évaluation managériale décrite plus haut. Enfin,

les évaluations dites de quatrième génération se sont développées sur le mode pluraliste autrement dit en associant les différents protagonistes à la démarche évaluative (Guba, Lincoln 1989).

● Démarches d'évaluation au sein de la Commission

Les débuts de l'évaluation à la Commission européenne remontent aux années 80 dans le domaine de la recherche, d'une part, et de l'aide aux pays en développement, d'autre part. La politique européenne de la science et de la technologie est fondée, depuis le début des années 80, sur un système de programmation pluriannuelle et l'évaluation a rapidement été intégrée dans la gestion de ces programmes. Un programme spécifique a d'ailleurs été mis en place (Monitor, 1989-1992) afin d'améliorer les méthodes et indicateurs dans le domaine de l'évaluation de la recherche. L'évaluation dans le domaine de l'aide aux pays en développement connaît également un essor significatif dans les années 80. L'approche communément utilisée est celle du *Project Cycle Management*. Elle est en partie reprise dans cet ouvrage (voir page 158). Les pratiques se sont ensuite étendues à d'autres domaines, notamment celui des Fonds structurels, avec la mise en place en 1988 d'une réglementation précise sur le suivi et l'évaluation des programmes.

En 1991, un peu plus de la moitié des Directions Générales disposaient d'une unité d'évaluation. Ce développement inégal peut s'expliquer par la genèse des politiques européennes. Certaines politiques disposent d'une légitimité *a priori*, autrement dit reconnue par les Traités comme de compétence exclusive des institutions européennes, par exemple la politique du commerce extérieur, celle de la concurrence, celle des marchés agricoles, etc. Ces politiques ont été peu évaluées jusqu'à une date récente. Les quelques évaluations existantes ont pour particularité d'être réalisées en interne. Ainsi, par exemple, dans le cas de la politique de la concurrence, les évaluations ont longtemps été faites par les fonctionnaires européens qui s'appuyaient sur la consultation des industriels concernés et des rencontres informelles avec les administrations nationales. La Commission, forte des textes juridiques qui définissent ses compétences dans ces domaines, peut conduire elle-même ses évaluations.

Par contre, dans d'autres domaines, la légitimité des interventions européennes s'est construite progressivement au fil des années. Il s'agit de politiques qui sont traditionnellement de compétence nationale telles que la recherche scientifique, la protection des consommateurs, les réseaux de

transports, etc. Malgré la reconnaissance croissante des initiatives européennes, ces politiques restent sujettes à justification. Les interventions sont reconnues comme nécessaires sous réserve que la Commission fasse la preuve des « avantages retirés de la conduite du programme à l'échelon communautaire ». Pour ce faire, elle passe commande de nombreuses évaluations. Ainsi, les évaluations des programmes de recherche s'efforcent de légitimer la politique européenne, en particulier vis-à-vis des Etats membres, en s'appuyant sur des comités d'experts réputés indépendants et recrutés dans les différents pays.

Dans le domaine de la politique de cohésion, l'évaluation a pris naissance en 1988 et s'est développée à un rythme très soutenu. Les évaluations sont parfois proches de la conception pluraliste décrite plus haut dans la mesure où certaines d'entre elles associent des responsables administratifs européens, nationaux et régionaux, ainsi que des partenaires socio-économiques, sans pour autant écarter les autres conceptions de l'évaluation. En outre, la Commission réalise de sa propre initiative (commande unilatérale) des évaluations thématiques ou des évaluations de programmes transversales à plusieurs pays. On trouvera dans la partie I de ce Volume, une analyse détaillée des pratiques d'évaluation des programmes des Fonds structurels.

Enfin, la Commission a lancé en 1995 un programme qui vise à améliorer sa gestion financière : « *Sound and Efficient Management - SEM 2000* ». Ce programme généralise l'évaluation à toutes les politiques ayant une incidence budgétaire. Il institutionnalise l'utilisation de l'évaluation selon le modèle managérial, comme élément clé de l'amélioration de la gestion au sein de la Commission. Les Directions Générales sont invitées à établir des programmes continus d'évaluation et à créer des unités d'évaluation. Des mécanismes de soutien ont également été mis en place (réseau des services d'évaluation, manuels, actions de formation, base de données des évaluations réalisées et projets pilotes). Tout renouvellement de programme pluriannuel est désormais conditionné par l'évaluation de ses résultats antérieurs.

■ Trois bonnes raisons d'évaluer

● Vérifier que l'action publique répond à des besoins non couverts ou insuffisamment satisfaits

Une des finalités de l'évaluation est de vérifier la raison d'être de l'action publique. En effet, une intervention publique ne se justifie que si elle est

destinée à satisfaire des besoins ou à résoudre des problèmes socio-économiques. L'évaluation sert à clarifier ces besoins ou à vérifier l'existence réelle de ces problèmes surtout si elle est ex ante, mais aussi lorsqu'elle est ex post pour vérifier si les besoins ou problèmes continuent d'exister.

De plus, il est impératif de statuer sur l'opportunité de l'intervention publique en tant que telle. Un programme public ne se justifie que si les besoins ne peuvent être satisfaits ni par l'initiative privée, ni par d'autres initiatives publiques déjà mises en place. Ainsi, le caractère insuffisant ou défaillant de l'initiative privée constitue la justification première de l'intervention publique en vertu du principe de subsidiarité, dans la définition originelle du terme qui veut que l'autorité publique agisse seulement quand le secteur privé est défaillant.

Pour ce qui concerne plus particulièrement les institutions européennes, ce concept de subsidiarité prend un sens supplémentaire dans la mesure où la légitimité de l'Union européenne doit également se justifier par rapport aux souverainetés nationales. Le Traité de Maastricht réaffirme l'importance du respect de la subsidiarité par rapport à l'échelon national : « la Communauté agit dans la mesure où un objectif peut être mieux réalisé au niveau européen qu'au niveau des Etats membres pris isolément ».

● Améliorer les interventions

Grâce à l'observation de la réalité des résultats et des impacts, l'exercice d'évaluation permet une rétroaction sur les actions en cours. Certaines leçons tirées de l'expérience débouchent sur des propositions d'améliorations dans la mise en œuvre sans nécessairement changer les hypothèses de départ. Par exemple, on diminuera la dotation budgétaire des actions qui réussissent le moins bien. Il s'agit alors d'une rétroaction directe ou d'un apprentissage simple (dit « en simple boucle »). L'évaluation peut également proposer des réorientations d'ordre stratégique, par exemple en réorganisant la mise en œuvre des actions pour qu'elles réussissent mieux. L'évaluation exerce alors une rétroaction indirecte ou un apprentissage en « double boucle » dans la mesure où elle remet en cause la logique de l'intervention.

Cette finalité de l'évaluation relève de ce qu'il est convenu d'appeler l'apport instrumental, autrement dit, d'une aide concrète en termes de solutions aux difficultés rencontrées ou d'adaptation à un environnement instable. Sa

fonction est qualifiée d'endoformative dans le sens où elle a comme vertu d'apporter tous les éléments nécessaires aux responsables opérationnels pour l'amélioration du programme : répartition plus rationnelle des ressources humaines et financières, amélioration de la gestion des services chargés de les mettre en œuvre, etc. Le but est d'analyser les facteurs de succès ou d'échec afin d'aider les gestionnaires à corriger les modalités d'intervention et de mise en œuvre du programme.

● **Rendre des comptes**

La troisième finalité de l'évaluation, consiste à rendre compte aux responsables politiques et aux citoyens des résultats obtenus et du bon emploi des ressources affectées. Cette finalité est désignée par les termes de responsabilité ou redevabilité, quelquefois d'imputabilité (*accountability*).

Les devoirs des autorités publiques en la matière sont la contrepartie du droit de lever l'impôt et du pouvoir de coercition. L'évaluation prend alors toute sa dimension politique dans le fonctionnement démocratique. La Commission, bien qu'elle ne soit pas responsable politiquement devant les électeurs européens, a affirmé progressivement une exigence de transparence vis-à-vis des contribuables sur la bonne utilisation de ces fonds qu'elle a la charge de gérer.

L'évaluation doit contribuer à expliquer de façon simple où a été l'argent public, quels effets il a produit et en quoi les dépenses étaient justifiées. Les destinataires de ce type d'évaluation sont bien évidemment les responsables politiques et *in fine*, le plus souvent par l'intermédiaire des médias, les citoyens. Une notion essentielle dans le cadre des évaluations européennes est celle de « valeur ajoutée communautaire » autrement dit le concours apporté par l'intervention des institutions européennes est-il justifié par des progrès réalisés ?

La spécificité des programmes et de leurs évaluations

Les programmes de développement socio-économique doivent être évalués à l'aide de méthodes adaptées à leurs spécificités. Une de leurs caractéristiques est d'être cofinancés et de faire l'objet de co-décisions. Ceci implique que l'évaluation soit conduite avec la participation de plusieurs partenaires et réponde à leurs différentes attentes.

Ces programmes sont aussi caractérisés par de multiples objectifs et domaines d'intervention. Il en résulte une plus grande complexité des travaux d'évaluation.

Enfin, contrairement à des contextes où l'évaluation est considérée comme un exercice purement rétrospectif, le système choisi pour la politique européenne de cohésion comporte une évaluation des programmes à trois stades successifs : ex ante, mi-parcours et ex post.

■ L'évaluation d'un ensemble d'objectifs combinés

Dès son origine, la politique européenne de cohésion économique et sociale a eu pour vocation la réduction des disparités économiques et sociales entre les régions d'Europe et leurs citoyens. Au début des années 1990, les écarts de développement et de niveau de vie restent marqués entre régions à l'échelle européenne. Les dix régions les plus pauvres de l'Union européenne avaient encore un produit intérieur brut par habitant plus de 3,5 fois inférieur à celui des plus riches. Le taux de chômage des dix régions les moins favorisées en matière d'emploi dépassait jusqu'à sept fois celui des dix régions les mieux placées.

La réduction de ces disparités, qui ne peut être le seul fait du libre jeu du marché, est un défi qui relève d'interventions publiques à très long terme. Les interventions de l'Union européenne viennent en appui de celles de chaque Etat membre en contribuant à la promotion d'un effort de solidarité européenne.

L'objectif de la cohésion économique et sociale a été introduit en 1986 dans l'Acte unique européen et a engendré la réforme de 1988 dont une augmentation considérable des ressources allouées à ses instruments financiers, les Fonds structurels : FEDER, FSE, FEOGA et plus récemment IFOP. Le Traité sur l'Union européenne, ratifié en 1993, a conforté cette orientation puisqu'il place la cohésion économique et sociale parmi les

objectifs essentiels de l'Union, avec en toile de fond, l'objectif de « convergence économique » indispensable à la réalisation du Marché unique mais aussi de l'Union économique et monétaire.

L'objectif global à long terme de cohésion de l'Union européenne se décline en plusieurs objectifs généraux tels que la solidarité et le soutien mutuel qui sont, à leur tour, exprimés en termes économiques et sociaux concrets et mesurables. La cohésion économique et sociale est officiellement définie dans l'article 130a du Traité sur l'Union européenne en termes de « développement harmonieux » avec une dimension géographique spécifique : les régions les moins favorisées, doivent pouvoir rattraper leur retard de développement. De plus, cette ambition de cohésion inclut le soutien aux groupes sociaux les plus défavorisés et plus particulièrement les chômeurs et les travailleurs dont l'emploi est menacé.

Agissant ensemble, les Fonds structurels sont, jusqu'en 1999, concentrés sur sept objectifs. Les objectifs 1, 2, 5b et 6 spécifiquement régionaux, concernent respectivement l'ajustement structurel des régions en retard de développement (Objectif 1), les reconversions économiques des zones industrielles en déclin (Objectif 2), la diversification économique des zones rurales vulnérables (Objectif 5b) et l'ajustement structurel des régions à très faible densité de population (Objectif 6). Les objectifs 3, 4 et 5a couvrant toute la communauté visent respectivement la lutte contre le chômage de longue durée, l'intégration professionnelle des jeunes et des personnes menacées d'exclusion du marché du travail (Objectif 3), l'adaptation des travailleurs aux mutations de l'industrie et des systèmes de production (Objectif 4) et l'adaptation des structures agricoles et de la pêche (Objectif 5a).

Afin d'accroître la visibilité et l'efficacité des Fonds structurels, le Conseil sur proposition de la Commission a décidé de ramener ces sept objectifs à trois : deux régionaux (régions en retard de développement et régions en reconversion économique et sociale) et le troisième horizontal consacré aux ressources humaines.

L'objectif global de réduction des disparités et de cohésion s'inscrit en outre dans le cadre de plusieurs priorités définies par les Etats membres et le Conseil qui doivent être respectées dans le cadre des objectifs de chacun des programmes et, *in fine*, dans les critères de sélection des projets financés. Ces priorités correspondent aux quatre enjeux transversaux que sont : la protection de l'environnement, la compétitivité, la création d'emplois et l'égalité des chances entre les femmes et les

hommes. Le Volume 5 de la Collection est consacré à la prise en compte de ces priorités dans l'évaluation.

En conséquence, l'évaluation des programmes doit s'attacher à l'analyse simultanée de plusieurs objectifs à long terme, d'un nombre encore plus grand d'objectifs à moyen terme, sans oublier de fournir des informations sur les mérites du programme par rapport aux priorités politiques transversales ci-dessus. Les responsables administratifs et plus encore les équipes d'évaluation doivent faire face à des difficultés techniques considérables du fait de cette imbrication d'objectifs qui se traduisent en impacts multiples et difficiles à observer et analyser en détail.

■ L'évaluation d'interventions dans des domaines multiples

Avec le renforcement du processus d'intégration européenne, l'étendue des domaines pour lesquels les Etats membres ont décidé de poursuivre une politique commune s'est accrue. Certaines sont spécifiquement ciblées sur l'amélioration de la cohésion et d'autres sont sensées y contribuer tout en ayant leurs objectifs propres.

Les interventions peuvent être regroupées en différentes rubriques : les interventions dans les secteurs agricole et de la pêche, les mesures visant à accroître la compétitivité (recherche et développement, politiques industrielle et commerciale), les politiques des réseaux (transports, télécommunication, énergie) ; les mesures visant à améliorer la qualité de vie (politique sociale, éducation et formation, protection et amélioration de l'environnement) ; etc.

Elles sont globalement couvertes par les quatre Fonds structurels (FEDER, FSE, FEOGA-Orientation, IFOP) :

- Le Fonds européen de développement régional (FEDER), créé en 1975, qui vise à renforcer le potentiel économique des régions aidées, à soutenir l'ajustement structurel et à contribuer à promouvoir la croissance des emplois durables. Il soutient l'investissement productif, des projets d'infrastructures et des actions de développement du potentiel économique propre des régions par le biais d'un cofinancement de programmes opérationnels, d'une aide aux grands projets, de subventions globales, d'une assistance technique et d'études préparatoires.
- Le Fonds social européen (FSE), créé en 1958 avec le Traité de Rome, mais réformé en 1988, qui a pour principal objectif de combattre le

chômage de longue durée et d'améliorer la capacité des jeunes à tenir un emploi. Depuis 1993 il s'y ajoute l'encouragement de l'adaptation aux mutations industrielles. Il contribue à financer la formation professionnelle et les mesures de soutien de l'emploi ainsi que les améliorations du système éducatif.

- Le Fonds européen d'orientation et de garantie agricole, partie « orientation » (FEOGA-O), créé en 1962 avec la Politique agricole commune, qui a pour objet d'encourager l'ajustement structurel dans l'agriculture grâce à des mesures de modernisation de la production et de développement des régions rurales.
- Et finalement, l'Instrument financier d'orientation de la pêche (IFOP) - créé en 1994 pour remplacer divers instruments financiers séparés en vigueur depuis 1976 - qui soutient l'adaptation des structures de ce secteur.

Le Fonds de cohésion et la Banque européenne d'investissement contribuent également à la politique de cohésion par des subventions et prêts sur projets. Le Fonds de cohésion, institué par le Traité de Maastricht, est destiné à faciliter la préparation de la Grèce, du Portugal, de l'Irlande et de l'Espagne à l'Union économique et monétaire. Ce Fonds peut être considéré comme une aide aux Etats membres les plus faibles en contrepartie du fait qu'ils s'engagent à satisfaire aux objectifs de convergence et à maîtriser leurs déficits budgétaires.

Dans le cadre de cette collection, et notamment dans ce Volume, on aura recours à une typologie d'intervention en trois grands domaines d'intervention couverts par les politiques mises en œuvre dans le cadre des Fonds structurels et du Fonds de cohésion, à savoir : l'amélioration des infrastructures de base (transport, communication, équipements publics en général), le développement des ressources humaines (par des actions de formation, de conseil auprès des entreprises et des individus) et le soutien à l'environnement productif comprenant les aides aux acteurs économiques pour, par exemple, accroître leurs capacités de production, d'exportation, d'innovation, etc. (voir page 147 et suivantes).

Cet éventail large de domaines hétérogènes d'intervention est fréquent en matière de programmes socio-économiques. Toutefois, cette diversité atteint, pour les programmes financés par les Fonds structurels, une ampleur rarement égalée. L'évaluation n'en est que plus délicate ; c'est pourquoi deux Volumes de la Collection traitent précisément des solutions méthodologiques possibles du point de vue des indicateurs (Volume 2) et des méthodes d'évaluation (Volume 4).

■ Plusieurs niveaux et plusieurs étapes d'évaluation

Dans le domaine public, l'évaluation est une opération qui peut concerner aussi bien des activités que des services, des organismes ou des pratiques. L'évaluation intervient également à trois niveaux principaux : projet, programme ou politique. Dans les pays anglophones, comme à la Commission européenne, c'est la notion d'évaluation de programme qui domine. Dans les pays latins, la littérature relative à l'évaluation fait plus souvent référence aux politiques publiques.

Le terme de « programme » désigne (1) un ensemble coordonné d'actions de différentes natures (par exemple des stages de formation, du conseil personnalisé, des subventions aux employeurs, etc.), (2) précisément défini dans ses objectifs opérationnels (ex. : on parlera d'un programme d'aide à la reconversion de chômeurs de longue durée), mais pas complètement défini à l'avance dans sa mise en œuvre et, enfin, (3) limité dans le temps et dans ses ressources budgétaires.

La notion de « politique », quant à elle, désigne un ensemble de programmes et/ou de mesures qui concourent à un même objectif général mais qui comportent des différences en termes d'objectifs spécifiques, de calendrier et de modalités de gestion (ex : une politique de l'emploi va se référer à un ensemble de mesures, pas toujours coordonnées, qui se sont accumulées au cours des années).

La définition des deux termes précédents renvoie à celle du « projet » qui est, pour sa part, une action non divisible et assortie d'un mode de gestion, d'un calendrier et d'un budget complètement précisés dès le départ.

Les trois niveaux d'évaluation sont pertinents pour la politique européenne de cohésion et ils méritent d'être articulés entre eux (voir page 55). Toutefois, c'est le niveau de l'évaluation des programmes qui fait l'objet de l'activité la plus importante, en termes quantitatifs. La Collection MEANS porte essentiellement sur l'évaluation à ce niveau.

Dans la littérature relative à l'évaluation, on peut lire des opinions contradictoires à propos de la temporalité de l'évaluation. Certains auteurs considèrent que l'évaluation est seulement un exercice rétrospectif alors que d'autres lui donnent aussi une dimension prospective. Dans le contexte de la politique européenne de cohésion, l'évaluation est clairement définie comme une activité qui intervient avant, pendant et après l'action publique. Les trois étapes ou stades de l'évaluation d'un programme sont reliés à son

cycle de vie et comprennent l'évaluation ex ante, l'évaluation à mi-parcours et l'évaluation ex post. Compte tenu de la succession ininterrompue des cycles de programmation, il apparaît un recouvrement temporel entre les évaluations réalisées à différents stades et pour différents cycles. Ceci impose de relier soigneusement les différents exercices pour qu'ils se renforcent mutuellement et ne soient pas redondants (voir page 51).

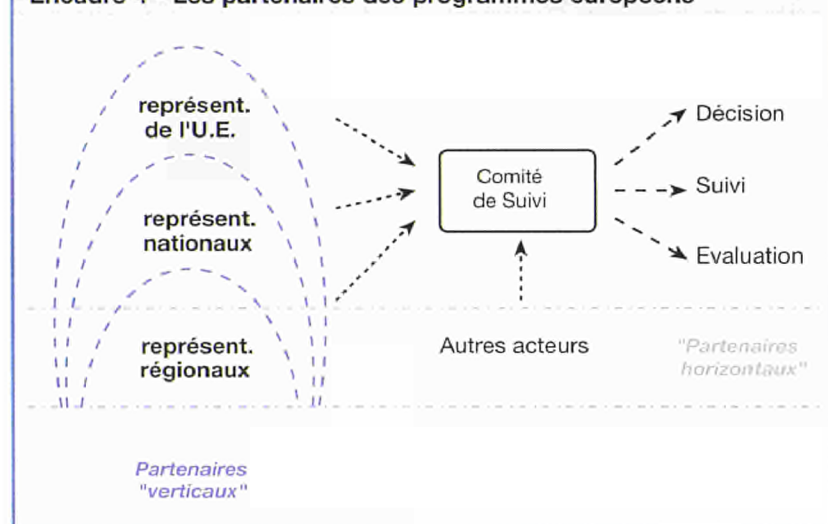
■ Partenariat et évaluation

● Le partenariat

Les programmes financés par les Fonds structurels sont systématiquement cofinancés par les Etats membres au niveau des autorités nationales et infra-nationales (Régions ou autres autorités locales). Le financement européen s'accompagne d'une exigence d'additionnalité. Son principe est d'assurer que l'aide communautaire au développement économique et social ne se substitue pas aux efforts des gouvernements nationaux, c'est-à-dire, que les financements propres de l'Etat bénéficiaire restent globalement au moins égaux à ceux qui existaient avant la contribution des Fonds structurels. Cela signifie concrètement que toute dépense communautaire vient en sus d'une dépense nationale. La vérification de la mise en œuvre de ce principe s'effectue à l'échelon national dans le cadre du contrôle financier et non de l'évaluation à proprement parler. Cette vérification présente encore des difficultés sur le plan méthodologique (données incomplètes, modalités de calcul, etc.).

La combinaison des financements européens et nationaux s'accompagne d'une conséquence décisive pour l'évaluation, à savoir l'existence de co-décisions et de partenariats. La conception et la mise en œuvre de programmes cofinancés nécessitent en effet la reconnaissance d'objectifs communs, l'établissement d'un cadre d'action contractuel et l'installation d'une confiance réciproque. Dans le cas des Fonds structurels, par gestion partenariale, on entend souvent le partenariat vertical, c'est-à-dire le fait qu'il doit s'établir une concertation étroite entre la Commission et toutes les autorités compétentes au niveau national, régional ou local désignées par chaque Etat membre. Il existe aussi plusieurs formes de partenariat horizontal : géographique, sectoriel, administratif, etc. Depuis 1994, le partenariat est étendu à tous les organismes compétents y compris (en règle générale) aux partenaires économiques et sociaux, désignés par l'Etat membre concerné.

Encadré 4 - Les partenaires des programmes européens



● L'évaluation partenariale

En termes d'évaluation, le partenariat, et tout particulièrement le partenariat « vertical » entre les différents niveaux d'administration, va engendrer des opportunités qu'il faudra saisir et des contraintes qu'il faudra dépasser. Le partenariat vertical représente une opportunité pour le développement d'évaluations pluralistes, dites de « quatrième génération » (voir page 22), mais introduit également toute une série de contraintes nouvelles. En effet, les politiques partenariales engendrent des situations qui rendent les évaluations difficiles à réaliser et qui en réduisent la portée. Les objectifs communs poursuivis manquent souvent de précision et de cohérence en raison des négociations et des compromis qui ont présidé à leur formulation. En outre, la coresponsabilité complique singulièrement la gestion des moyens. L'évaluation a donc des difficultés à définir son objet et fait craindre aux partenaires une remise en question du *modus operandi* négocié.

L'évaluation partenariale, devant tenir compte de la diversité des attentes propres à chaque partenaire et rechercher la convergence des points de vue, se trouve confrontée au fait qu'elle ne représente pas le même enjeu pour tous les partenaires. Souvent, l'existence d'enjeux spécifiques est gommée au profit du postulat que ce qui est bon pour un niveau de collectivité l'est forcément pour les autres. Cette hypothèse implicite a pour

conséquence l'absence de différenciation des questions posées et des critères de jugement. Finalement, une évaluation qui, sous prétexte de partenariat, serait construite en négligeant les différences initiales entre les partenaires court le risque de devenir une construction procédurale sans substance.

D'une situation partenariale à l'autre, il existe, *de facto*, de profondes différences dans l'application de l'obligation réglementaire d'évaluer. Dans un certain nombre de pays ou de régions, cette obligation se réduit à un exercice formel assez superficiel. Dans d'autres cas, elle a conduit non seulement à une meilleure efficacité des Fonds structurels mais aussi à une évolution générale des modèles administratifs. Ainsi, par exemple, dans plusieurs Etats membres et régions récemment éligibles, les exigences communautaires en matière d'évaluation ont contribué à promouvoir de nouvelles pratiques dans le cadre de la programmation et de la mise en œuvre des politiques régionales.

On remarque par ailleurs que l'application de cette réglementation varie autant entre les différents Etats membres qu'entre régions au sein des Etats eux-mêmes. Il apparaît que les évaluations ont plus de chance d'être de qualité lorsque les relations entre les partenaires sont équilibrées, autrement dit lorsqu'un financeur n'est pas trop dominant. Les régions les plus riches ne sont pas nécessairement les plus respectueuses des règlements des Fonds structurels probablement parce qu'ils ne représentent pas un apport budgétaire considérable pour ces collectivités. Les régions ou les pays les moins favorisés, quant à eux, reçoivent des aides très importantes mais ne sont pas toujours les plus actives en matière d'évaluation, soit parce qu'ils rencontrent déjà des difficultés pour mettre en œuvre les ressources disponibles, ou encore parce qu'ils perçoivent l'évaluation comme une menace compte tenu de leur position de débiteurs vis-à-vis des autres partenaires. L'hypothèse peut être faite que la volonté d'évaluer de chaque partenaire dépend de son appréciation des gains mais aussi des coûts et des risques liés à l'évaluation de son programme.

A l'avenir, l'action de la Commission européenne sera davantage centrée sur le choix des priorités stratégiques, laissant aux autorités nationales, régionales et locales le soin de piloter la programmation et la gestion des programmes. Cette gestion plus décentralisée et simplifiée aura pour contrepartie un suivi et une évaluation renforcés.

Le développement des pratiques dans le cadre des Fonds structurels européens

La diffusion de l'évaluation dans le cadre de la politique européenne de cohésion date de moins de dix ans. En s'étendant dans les Etats membres, elle rencontre des références et des schémas de pensée très divers (expertise macroéconomique, économie régionale, analyse politique, analyse stratégique, etc.) et se fonde sur des concepts qui peuvent varier d'un pays à l'autre.

Néanmoins, quels que soient le vocabulaire et les structures adoptées, les systèmes d'évaluation mis en place dans les différents Etats membres ont contribué à accroître la transparence et l'efficacité des interventions financées par les Fonds structurels, et souvent même, d'autres interventions publiques dans des domaines proches.

A partir d'une pratique quasi inexistante, des progrès considérables ont été faits dans la qualité des rapports d'évaluation. Il subsiste néanmoins un certain nombre de difficultés, en particulier dans le domaine de l'analyse des impacts et dans celui de l'utilisation des rapports.

■ L'évolution de la réglementation des Fonds structurels

Au niveau de la Commission une exigence de transparence vis-à-vis des contribuables des Etats membres s'est affirmée progressivement, sur la bonne utilisation des fonds d'origine nationale qu'elle a la charge de gérer. Cette aspiration à un bon usage des deniers européens s'est traduite par une expression simple : « recherche de l'efficacité des dépenses européennes ».

L'Acte unique européen, adopté au début 1986, a introduit au Traité CEE un nouveau titre V contenant un article 130D qui annonçait la réforme des Fonds structurels, destinée à améliorer leur efficacité en vue d'une meilleure cohésion économique et sociale de la Communauté. Dans le même esprit, en février 1992, le Traité de Maastricht a repris dans son titre XVI cet impératif de cohésion et l'article 130B stipule que le Conseil définit les dispositions nécessaires pour assurer l'efficacité des Fonds.

Conformément à ces impératifs, les règlements relatifs aux Fonds structurels (ceux de 1988 pour la première génération des Fonds structurels et ceux de 1993 pour la génération actuelle) ont été pourvus d'articles

spécifiques consacrés à l'évaluation ; notamment l'article 6 du règlement-cadre et l'article 26 du règlement de coordination.

● **La réglementation appliquée de 1988-93**

Dans son paragraphe 2, l'article 6 du règlement de 1988 stipulait ainsi que « en vue d'apprécier l'efficacité des interventions structurelles, l'action communautaire fait l'objet d'une évaluation ex ante et ex post destinée à apprécier son impact et que pour ce faire, les demandes de concours établies par les autorités compétentes devraient comporter les informations nécessaires pour pouvoir être évaluées par la Commission ».

Ces premiers règlements, adoptés par le Conseil le 19 décembre 1988, concernaient des plans de développement régional qui devaient être déposés le 1er janvier 1989. Pour un certain nombre d'administrations nationales ou régionales, la nécessité d'établir une programmation pluriannuelle constituait une sorte de révolution. Quant à la maîtrise de la démarche d'évaluation, elle était très réduite aussi bien de la part de la majorité des autorités compétentes que des services de la Commission eux-mêmes, au moins en ce qui concerne les Fonds structurels. Par conséquent, les évaluations réalisées pendant la période 1989-1993 n'ont pas été d'une très grande qualité malgré les efforts réalisés par un certain nombre d'acteurs, surtout au niveau de la Commission. On a observé notamment que le cadre défini par cette première version du règlement ne créait pas les conditions adéquates pour que la phase d'évaluation ex ante se déroule normalement.

● **La réglementation applicable en 1994-99**

Admettant ce demi-échec, à l'instigation de la Commission, mais également sous la forte impulsion de certains Etats membres (Pays-Bas et Royaume-Uni en particulier), les nouveaux règlements de juillet 1993 ont renforcé considérablement les exigences en matière d'évaluation.

Cette nouvelle réglementation (article 6 du règlement cadre) précise les notions d'appréciation ex ante, de suivi et d'évaluation ex post. Les articles détaillant le contenu des documents de programmation déposés par les Etats membres introduisent la notion « d'objectifs spécifiques quantifiés si leur nature s'y prête ». De même, l'article 26 du règlement de coordination présente une clause de conditionnalité spécifiant « que les aides seront octroyées lorsque l'appréciation ex ante aura démontré les avantages socio-économiques à retirer à moyen terme, eu égard aux ressources mobilisées ».

Cette exigence très forte d'évaluation ex ante, variante communautaire de la notion britannique de *Value For Money* n'a en pas été très strictement appliquée dans la pratique. Cependant l'introduction d'évaluations intermédiaires qui n'existaient pas jusqu'alors a rencontré un succès significatif comme le montre plus loin notre enquête. Il est clair que ce renforcement de la législation européenne en matière d'évaluation a constitué un stimulant puissant pour que se répande l'idée que l'obligation d'évaluer était un point de passage incontournable pour recevoir des Fonds communautaires.

● La réglementation de la troisième génération de programmes (2000-2006)

La troisième génération de Fonds structurels verra donc se renforcer l'appréciation de leur efficacité. Les actions communautaires feront désormais l'objet d'une évaluation ex ante par l'Etat membre lui-même, d'une évaluation à mi-parcours par les gestionnaires du programme et d'une évaluation ex post à l'initiative de la Commission, destinées à apprécier leurs effets par rapport aux objectifs (objectifs propres et objectifs de cohésion) et à analyser leurs incidences sur des problèmes structurels spécifiques.

Ainsi, l'efficacité de l'action des Fonds sera mesurée à trois niveaux : effet d'ensemble sur les objectifs visés à l'article 130 A du Traité (notamment le renforcement de la cohésion économique et sociale), effet sur les priorités proposées dans les Plans et prévues dans chaque Cadre Communautaire d'Appui, effet sur les priorités spécifiques retenues au titre des interventions. Des évaluations complémentaires, le cas échéant thématiques, pourront être lancées en vue d'identifier des expériences transférables d'un programme à l'autre.

Afin de renforcer la transparence, il est prévu que les rapports d'évaluation soient mis à la disposition du public plus largement que cela n'a été le cas jusqu'à présent.

■ L'évolution récente des pratiques d'évaluation¹

● Les appréciations ex ante : un constat en demi-teinte

L'évaluation de la deuxième génération de programmes, avec l'impulsion de la réglementation révisée et renforcée de 1993, dispose d'un cadre

¹ Cette section est fondée sur une enquête conduite en 1997 dans le cadre du programme MEANS. L'enquête a comporté des entretiens téléphoniques dans tous les Etats membres et des visites dans cinq pays.

rigoureux. Notamment, la phase d'appréciation ex ante doit préciser les objectifs au sein de la stratégie régionale et vérifier la prise en compte des priorités communautaires, fournissant ainsi aux services responsables des éléments pour juger de la pertinence des actions lors des négociations.

En règle générale, les appréciations ex ante sont conduites par les administrations nationales. Pour la première fois, elles l'ont été également de manière systématique par la Commission. Toutefois, les appréciations ex ante n'ont pas toujours eu l'utilité attendue. Pour des raisons de calendrier, les analyses sont souvent restées trop superficielles pour produire une valeur ajoutée, tant au niveau de la formulation des objectifs généraux et spécifiques des programmes que des modalités de mise en œuvre de ceux-ci.

Des efforts ont cependant été réalisés par les Etats membres, notamment au niveau de la quantification des objectifs. Incontestablement, les Documents de Programmation de la période 1994-1999 sont de meilleure qualité que ceux de la période précédente. Ils ont été enrichis grâce à des réflexions conduites dans le cadre des analyses prospectives qui ont contribué à clarifier les objectifs poursuivis, à structurer les composantes des programmes et à hiérarchiser les priorités.

● **Le développement spectaculaire des évaluations à mi-parcours**

Des progrès notables ont été réalisés dans le cadre des évaluations à mi-parcours du deuxième cycle de programmation des Fonds structurels. Cette période ayant bénéficié des acquis de la période précédente, les évaluations ont en général été systématiquement réalisées.

Sur le plan quantitatif, le nombre de rapports produits a été multiplié par cinq ou six par rapport à la période précédente et la pratique d'évaluation s'est développée de façon spectaculaire.

Ainsi, dans les quinze Etats membres, pour la période de 1996 au début de 1998, plus de 380 évaluations à mi-parcours ont été organisées par des Comités de suivi pour les six Objectifs. Parmi celles-ci, on ne compte pas moins de 250 évaluations pour les seuls Objectifs 1 et 2 et pour l'année 1997 uniquement. A titre comparatif, pour la période 1989-1993, on a réalisé en moyenne moins de 40 évaluations par an sur l'ensemble des programmes et pour toutes les formes d'évaluation (ex ante, ex post, thématiques). Le nombre d'évaluations de programme dans le cadre des Fonds structurels a donc été multiplié par 5 ou 6 d'une génération à l'autre.

Des évaluations systématiques globales et parfois des analyses approfondies thématiques et géographiques ont été réalisées pour les programmes Objectif 1 et 5b. Ces évaluations à mi-parcours visent à fournir les premiers résultats du programme et indiquer les ajustements à effectuer en cours de programme 1994-1999.

Pour les programmes Objectif 2, les évaluations ont été conduites en 1996-97, à la fin de la première phase de programmation (1994-1996) dans la majorité des pays. Les nouveaux Etats membres ont également fourni leurs rapports d'évaluation. Compte tenu du lancement tardif des programmes, la période de mise en œuvre prise en compte dans les rapports a été plus courte. Il n'en demeure pas moins que les travaux ont été utiles. En Suède par exemple, l'évaluation des programmes Objectif 2 a été reconnue comme utile pour dégager les premières conclusions sur les dispositions de mise en œuvre et l'adéquation des mesures.

On constate généralement que l'utilité de l'exercice du point de vue managérial s'est renforcée au fil des expériences au point que de nombreux gestionnaires sont allés au-delà des exigences réglementaires. En plus des évaluations thématiques et transnationales faites à l'initiative des services de la Commission, de nombreuses évaluations sectorielles ont été lancées par des administrations nationales ou des gestionnaires de programme.

● Une amélioration notable de la qualité des rapports d'évaluation

La qualité moyenne des évaluations réalisées au cours de la période 1994-1999 s'est incontestablement accrue par rapport à la période précédente, notamment pour ce qui est des évaluations à mi-parcours. Il reste certes des travaux de qualité médiocre qui peuvent s'expliquer par les retards pris dans le lancement de certains programmes dans certaines régions d'Objectif 1 ou d'Objectif 5b.

Une critique commune faite à la plupart des rapports est leur caractère peu analytique. Les conclusions restent trop descriptives pour apporter une aide efficace à la prise de décision. Il existe néanmoins des pratiques correspondant aux meilleurs standards professionnels. Ainsi des évaluations (notamment en Belgique et en Irlande) ont su fournir des conclusions crédibles sur la question délicate des réallocations budgétaires : identification de mesures à renforcer ou à reconsidérer.

La plus grande expérience des chargés d'évaluation dans le domaine et une meilleure connaissance de la spécificité de la démarche d'évaluation ont permis de produire de meilleurs rapports. Beaucoup d'entre eux sont construits pour l'essentiel sur l'analyse de données secondaires (collectées à partir de bases de données ou de documents existants) et des indicateurs de suivi qui se sont améliorés au cours des dernières années. Sur la base de ces données, les Etats membres ont ainsi mieux répondu aux exigences de la réglementation qui imposent l'examen des données physiques et financières dans le cadre de l'évaluation. D'importants travaux ont été réalisés par exemple en Italie où un système d'indicateurs standards dans le domaine de la formation a été établi et au Portugal où des propositions d'indicateurs ont été construites pour l'ensemble des secteurs.

C'est au niveau de la collecte et du traitement des données socio-économiques que les pratiques nécessitent le plus d'améliorations. En effet, les données primaires, collectées directement par enquêtes auprès des publics concernés, restent rares et la qualité des informations fournies par les indicateurs de contexte ne tire pas encore tout le parti possible des systèmes statistiques nationaux et régionaux. Certains de ces systèmes statistiques sont parfois peu développés (cas de la Grèce et du Portugal par exemple) ou, comme dans le cas des nouveaux Länder allemands, insuffisamment harmonisés avec les systèmes européens (EUROSTAT). C'est dans cette perspective de tirer un meilleur parti des données existantes qu'a été écrit le Volume 2 de cette Collection.

● **La multiplication des initiatives et le renforcement des moyens**

Les services de la Commission ont fourni des efforts remarquables en matière d'évaluation, par eux-mêmes mais aussi en étroite coopération avec les administrations nationales.

A l'actif des services communautaires il faut évoquer, à partir de 1994, la création ou le renforcement d'unités d'évaluation et la diffusion des compétences spécialisées dans l'ensemble des services opérationnels. Entre 1994 et 1998 près de 260 fonctionnaires en charge des Fonds structurels ont suivi une formation à l'évaluation de deux ou trois jours dans le cadre du programme MEANS et de SEM 2000 (soit pratiquement la moitié des fonctionnaires concernés).

Selon les Etats membres, les administrations nationales ont un rôle primordial à l'échelon central (Department of Finance en Irlande, Ministères

de l'Agriculture, des Transports et Chancellerie en Allemagne) ou « déconcentré » (Préfectures en France). Dans certains pays les responsabilités sont assumées par les deux types d'autorités (nationales et régionales), comme en Espagne. Aux Pays Bas par exemple, les autorités nationales ont élaboré un cadre méthodologique (voir exemple ci-dessous) afin de garantir la pertinence et la qualité des rapports d'évaluation réalisés sous la responsabilité des différents comités.

Exemple : Le cadre méthodologique néerlandais

Ce pays a développé en 1994 un cadre méthodologique commun, appliqué à toutes les étapes d'évaluation et pour l'ensemble des programmes des Fonds structurels. Chaque rapport d'évaluation est ainsi construit de la même manière : situation socio-économique de la région, stratégie et cohérence, organisation et suivi, mise en œuvre et synergie, effets. Ce cadre, associé à l'utilisation d'indicateurs communs aux différentes régions et à un système de suivi performant, permet un échange d'informations et d'expériences, du fait de la comparabilité des outils utilisés dans les régions concernées par les programmes des Fonds structurels.

La participation des autorités régionales aux exercices d'évaluation s'est sensiblement renforcée depuis 1994 : programmation, définition, sélection des évaluateurs, prise en compte des recommandations, etc. Ce renforcement peut être considéré comme un facteur de développement de l'évaluation surtout lorsque ces autorités régionales disposent d'une autonomie forte et constituent par conséquent un partenaire dominant.

En Belgique, le niveau fédéral n'intervient pas dans la gestion des programmes structurels et l'évaluation est par conséquent gérée au niveau des régions. En Allemagne, les administrations des Länder sont les principaux responsables de la mise en œuvre des Fonds structurels et donc de leur évaluation. Il subsiste toutefois des différences importantes entre les anciens et les nouveaux Länder allemands. Les premiers, habitués à la démarche évaluative, n'impliquent le niveau national que lorsque c'est indispensable, alors que les seconds sont systématiquement soutenus par les administrations nationales.

Dans des Etats comme le Portugal ou la France, les Ministères techniques ne sont pas forcément très actifs dans l'évaluation des Fonds structurels. En revanche, les administrations coordinatrices au niveau national assurent un rôle clef d'animation et de suivi permanent des travaux d'évaluation conduits sous la responsabilité de chacun des Comités de suivi.

■ La dynamique créée par les Fonds structurels

● La valeur ajoutée du partenariat

Le partenariat, en créant des relations de confiance, a crédibilisé l'évaluation et en a ainsi facilité la mise en œuvre. La réserve, voire l'hostilité, qui pouvait persister vis-à-vis de la démarche depuis le début des années 90 s'est atténuée pour disparaître dans la majorité des Etats membres ; ce qui a bien entendu permis de mettre en œuvre l'évaluation beaucoup plus facilement.

La gestion des programmes des Fonds structurels implique de nombreux acteurs et une coordination entre les niveaux européen, national et régional. Le partenariat s'est construit dans le cadre des Comités de suivi, incluant les acteurs socio-économiques. Pour ce qui est de l'évaluation, les Comités de suivi ont, dans pratiquement tous les pays, confié la responsabilité de la démarche à des Comités de pilotage (ou comités techniques, comités d'accompagnement, etc.) de taille plus restreinte, donc mieux à même d'aborder des questions techniques. Ces comités sont le plus souvent composés d'autorités européennes, nationales, régionales et parfois d'experts (universitaires, statisticiens, etc.). Ils jouent un rôle primordial dans la conception méthodologique et l'examen de la qualité des travaux d'évaluation. Les rapports d'évaluation sont discutés au sein de ces groupes restreints avant d'être présentés aux Comités de suivi en séance plénière. L'application de cette organisation au Portugal est présentée ci-dessous.

Exemple : Le partenariat, un facteur clé de succès au Portugal

Chaque Comité de suivi constitue un groupe technique d'évaluation comprenant les rapporteurs de la Commission, des représentants des administrations nationales et régionales portugaises et chargé de gérer les relations avec l'équipe d'évaluation. Ce groupe technique se réunit au moins trois fois par an (pour valider les cahiers des charges, sélectionner les offres, discuter les rapports avant remise au Comité de suivi). Les bonnes relations entre partenaires au sein des groupes techniques d'une part et, d'autre part, les interactions constructives entre les commanditaires et les équipes d'évaluation constituent deux points forts qui ont eu, de l'avis des responsables nationaux, des effets très positifs sur la qualité des évaluations.

Le partenariat avec les services de la Commission a joué également un rôle positif : apports en terme de documentation technique aux différentes phases, participation à la sélection des équipes d'évaluation ou encore analyse de la qualité des rapports.

De nombreuses initiatives ont été lancées par la Commission afin de diffuser les connaissances et promouvoir les échanges entre Bruxelles et ses partenaires. La Commission a mis en place des Groupes Technique d'Evaluations (comprenant des représentants des autorités nationales et des services de la Commission) qui se réunissent plusieurs fois par an. Au niveau de la commande d'évaluation, la Commission a élaboré en partenariat avec plusieurs Etats des modèles de cahiers des charges. Des réunions d'information et d'échange entre la Commission et les gestionnaires de programmes ont lieu régulièrement au sein des Etats membres (en Grèce et en Finlande par exemple).

On ne peut toutefois aborder cette question du partenariat dans le cadre des Fonds structurels sans évoquer une des conclusions produites par le programme MEANS en 1996 et qui reste d'actualité, à savoir que si le partenariat est un moteur du développement des pratiques, il peut, à cause de trop nombreux compromis entre les co-décideurs, engendrer une moindre qualité des évaluations.

● La professionnalisation des équipes d'évaluation

Le marché de l'évaluation au cours des cinq dernières années a connu un développement remarquable, tout comme le nombre de structures réalisant des travaux d'évaluation. L'offre de service en évaluation reste le plus souvent extérieure aux administrations à l'exception cependant des administrations centrales irlandaises, françaises et italiennes qui ont créé en leur sein des équipes spécialisées. Lorsqu'elles existent, ces unités d'évaluations dont l'indépendance est garantie par un statut juridique spécial et par une position « hors hiérarchie », bénéficient d'une expérience que les praticiens externes n'ont pas forcément car l'évaluation ne représente qu'une part de leurs activités professionnelles. Le cas de l'administration irlandaise est présenté en page suivante.

Le nombre de cabinets de consultants actifs dans le domaine de l'évaluation s'est très fortement accru dans l'ensemble des pays. Le fait de confier l'évaluation à des évaluateurs externes, qu'ils soient consultants privés ou universitaires, est considéré comme une assurance d'indépendance de l'équipe par rapport aux autorités responsables. Le secteur d'activité s'est développé selon différents modes : émergence d'équipes liées au monde académique ou, au contraire, création de départements ou filiales de grands groupes internationaux de consultance. La Grèce et le Portugal ont ainsi vu

se développer un nombre important de petites structures (plus d'une trentaine de cabinets sont actuellement impliqués dans la réalisation d'évaluations des programmes structurels). On constate parfois, comme c'est le cas en France, un phénomène de concentration : quelques cabinets se partagent l'essentiel du marché.

Exemple : Des équipes d'évaluation permanentes au sein des administrations irlandaises

Le gouvernement irlandais et la Commission ont établi dès 1992 des unités spécifiques d'évaluation (Evaluation Units) dans les domaines de l'Industrie et de l'Agriculture, ainsi qu'une unité spécifique pour le FSE et entre autres pour l'ensemble du CCA, qui joue un rôle « horizontal » de coordination. La mise en place de ces unités, par leur indépendance et leur professionnalisme, contribue au bon déroulement de l'évaluation et à la qualité des travaux produits.

Les unités entreprennent des évaluations thématiques et des analyses approfondies. Leur programme de travail est décidé par des Comités de pilotage pluralistes. La plupart des évaluations sont conduites par les unités elles-mêmes, l'unité d'Agriculture faisant parfois appel à des compétences extérieures. Toutes ces évaluations sont publiées.

Les conditions pour réaliser des travaux de qualité ne sont toutefois pas toujours réunies. Ainsi par exemple, les délais sont souvent trop courts (un à six mois) pour des raisons de calendrier. On trouve parfois des délais plus raisonnables (allant jusqu'à douze mois) et dans certains pays comme la Finlande, l'Irlande, la France ou l'Allemagne ou encore lorsqu'il s'agit d'évaluations approfondies. En corollaire, les budgets alloués sont faibles (30 KEuros pour certaines évaluations aux Pays-Bas et en Espagne) pour pouvoir réaliser de véritables enquêtes de terrain auprès des individus ou des entreprises concernées.

On constate que le coût des travaux d'évaluation reste en règle générale très modeste si on le rapporte au budget global du programme : 0,02 % en moyenne et probablement moins de 0,01% dans les régions d'Objectif 1. Toutefois, ces ratios sont amenés à se rapprocher des normes internationales (voir page 83) lorsque seront connus les budgets consacrés aux évaluations approfondies, en cours et ex post, à venir.

Le développement quantitatif de l'offre s'est également accompagné d'un développement qualitatif engendré par plus de professionnalisme. Les compétences ont en effet évolué au cours des dernières années, davantage

par l'expérience accumulée que par une spécialisation préalable des prestataires dans le domaine de l'évaluation. Les équipes d'évaluation ont souvent des compétences pluridisciplinaires même si les économistes jouent un rôle dominant dans le cadre des évaluations des programmes régionalisés.

Les consultants font de plus en plus appel à la collaboration d'universitaires qui disposent d'une expertise plus pointue et précise dans tel ou tel domaine des politiques publiques. Nombreux sont par ailleurs les bureaux d'études créés par des universitaires, surtout dans les pays du sud de l'Europe (Portugal, Espagne, Italie). Les compétences en Allemagne, Finlande et Belgique reposent majoritairement sur les universités, les instituts ou les centres de recherche.

Certaines avancées méthodologiques réalisées au cours des dernières années méritent d'être soulignées. Les savoir-faire en modélisation de l'économie régionale se sont ainsi accrus en Allemagne, Italie, Irlande et plus récemment en Belgique. Des analyses *Input-Output* ont été mises en œuvre avec compétence pour la conduite d'évaluations au Portugal, en Allemagne, en Italie et en Espagne. L'analyse des relations de causalité s'est développée et les spécialistes font davantage appel à des analyses statistiques comparatives. Encore limitée il y a quelques années au Danemark et au Royaume-Uni, la pratique des groupes de comparaison s'est diffusée en Suède, en Allemagne, en France ou encore en Espagne, principalement dans le domaine de l'emploi.

● **La construction de capacités d'évaluation au sein des Etats**

On a constaté que les spécificités institutionnelles de chaque Etat conduisent à des différences importantes du point de vue de la mise en œuvre de l'évaluation. Il en va de même pour ce qui est de l'appropriation et de l'utilisation des résultats des rapports par les commanditaires. Le paysage européen, en matière d'utilisation de l'évaluation, est, comme on peut l'imaginer, assez contrasté. Cependant, de manière un peu schématique, on peut discerner trois phases dans le niveau d'utilité attribué à l'évaluation ; et par là même, dans la qualité de son organisation.

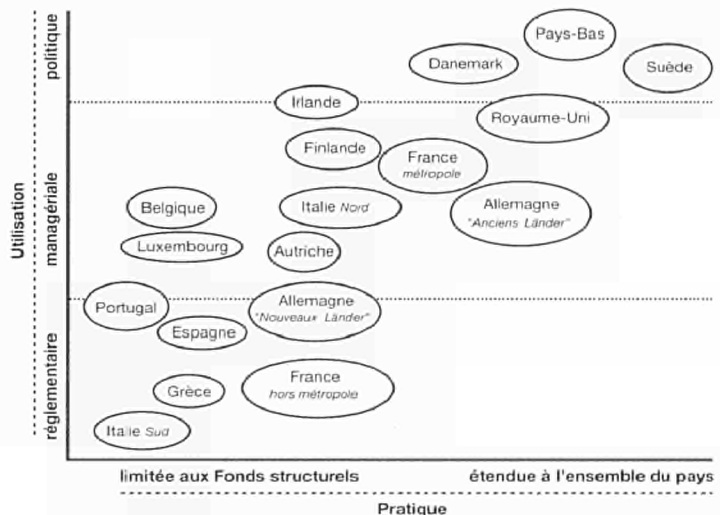
1ère phase : L'évaluation est vécue comme une réponse à des obligations réglementaires. C'est donc une contrainte et une charge de travail supplémentaire qui pèsent sur les gestionnaires qui considèrent que l'évaluation est avant tout exigée par Bruxelles ; même si les règlements sont des décisions du Conseil adoptées à l'unanimité.

2ème phase : L'évaluation devient un système d'aide à la conception et à la gestion des interventions. Un dialogue s'établit avec les évaluateurs et la qualité des informations recueillies s'améliore. Dans cette phase très opérationnelle, la démarche d'évaluation se raffine et les progrès sont rapides.

3ème phase : L'évaluation devient un acte politique et ses résultats font l'objet de débats publics. Le but recherché est d'informer l'opinion publique sur l'efficacité de l'utilisation des Fonds publics et sur la démonstration de leur efficience en termes d'obtention des impacts attendus (*Value For Money*). Dans un certain sens, l'évaluation devient un outil du jeu démocratique en informant sinon en recherchant le consentement du citoyen sur les dépenses contrôlées par l'Etat.

Ces trois phases sont représentées de manière simplifiée dans le schéma ci-dessous qui met en perspective les différents Etats membres, en fonction de leur utilisation des travaux d'évaluation, telle qu'elle vient d'être exposée. D'une part, l'utilisation de l'évaluation est représentée par l'axe vertical, et d'autre part, la pratique d'évaluation au sein des pays est représentée par l'axe horizontal. Concernant la pratique, on regarde si celle-ci se limite aux Fonds structurels ou si elle est également pratiquée dans le cadre d'autres politiques nationales ou régionales.

Encadré 5 - Pratique et utilisation de l'évaluation



Ce graphique n'est qu'une image instantanée (enquête réalisée en 1998). Il conviendrait de représenter également les dynamiques propres à chaque pays. Le Portugal, par exemple, s'est considérablement familiarisé avec l'exercice d'évaluation, et au cours des dernières années, d'importants progrès ont été réalisés en termes de qualité et d'utilisation des travaux.

L'influence de la réglementation communautaire sur le développement des pratiques est sensible dans la majorité des Etats membres. Ainsi, dans les pays d'Objectif 1, les travaux d'évaluation sont essentiellement ceux consacrés aux Fonds structurels. Cette influence, bien que perceptible, semble cependant moins marquée dans certains pays comme en Irlande, où les réformes récentes liées à l'amélioration de la gestion des actions publiques, émanent plutôt du mouvement de *performance management* développé dans les pays anglophones. Dans certains Etats, les exigences européennes régissant la mise en œuvre des politiques structurelles ont provoqué des changements d'attitude remarquables. Les obligations réglementaires ont conduit à une application systématique de l'exercice. Ainsi, par exemple, en Allemagne et en France, l'évaluation était en cours d'institutionnalisation mais la réglementation communautaire a accéléré le processus existant.

En ce qui concerne les pays les plus récemment entrés dans l'Union européenne, la réglementation communautaire a eu des effets d'autant plus notables que la pratique, les compétences et la culture d'évaluation étaient limitées. L'obligation d'évaluer, imposée par la Communauté, a provoqué l'émergence d'un nouveau marché accompagné d'une institutionnalisation accrue et d'une culture d'évaluation plus affirmée. On retrouve ce phénomène en Finlande où les pratiques se développent rapidement grâce à la relative intégration des politiques structurelles européennes aux politiques nationales finlandaises.

Il est intéressant de remarquer que l'utilisation démocratique de l'évaluation est favorisée lorsqu'il existe une pré-culture indépendamment des exigences communautaires. Autrement dit, l'obligation réglementaire peut aider à établir une utilisation managériale mais n'est probablement pas suffisante pour induire une contribution au débat public. A l'inverse, le nombre de pays figurant dans la partie inférieure gauche de l'encadré ci-dessus traduit l'importance de la réglementation communautaire sur le plan institutionnel et son influence sur le développement des pratiques dans des pays où la culture d'évaluation était auparavant faible.

Les réflexions internationales sur la construction de capacités d'évaluation insistent sur l'importance des actions de sensibilisation et d'incitation comme facteurs importants du développement de l'évaluation. Diverses initiatives peuvent y contribuer parmi lesquelles la création de sociétés ou d'associations nationales d'évaluation. La société allemande et l'association nationale italienne d'évaluation ont ainsi été récemment créées (Boyle, Lemaire, à paraître). La société britannique, particulièrement active, est brièvement présentée ci-dessous.

Exemple : La communauté professionnelle britannique de l'évaluation

Les équipes britanniques du secteur privé et des universités ont développé des compétences et une expertise dans le domaine de l'évaluation qui sont reconnues à l'échelle internationale. Le Royaume-Uni est un des rares Etats membres à disposer d'équipes de recherche en évaluation qui travaillent dans ce domaine de manière « transversale », sans se limiter à un aspect méthodologique particulier ni à un secteur politique donné. La Société d'évaluation britannique (UKES), forte de plus de 200 membres, organise une conférence annuelle et des ateliers de travail pour développer les échanges méthodologiques.

La formation est un outil essentiel de sensibilisation. Les pays dans lesquels les personnels administratifs ont bénéficié de formations spécifiques à l'évaluation restent rares. Cependant, là où ces activités ont pu être mises en place (séminaires sur l'évaluation organisés en France par des cabinets privés ou par les services administratifs, ou encore en Grèce, par les services de la Commission), les résultats sont positifs. On notera également que quelques cours sur l'évaluation des politiques publiques existent ou se développent en milieu universitaire, le plus souvent dans le cadre de formations sectorielles.

A partir du précédent constat relatif à l'évolution des pratiques au cours des dernières années, on peut identifier deux défis majeurs à relever pour soutenir la dynamique actuelle de développement de l'évaluation dans le cadre des Fonds structurels :

- renforcer le professionnalisme des praticiens pour compléter des compétences insuffisantes, tant au niveau des responsables administratifs que des équipes d'évaluation et renforcer la qualité des travaux ;
- mettre en œuvre des actions de sensibilisation et d'incitation efficaces, autrement dit passer d'une réponse à une obligation à une démarche volontaire.

II

Préparer l'évaluation et définir son contenu



Définir le cadre général de l'évaluation projetée

L'évaluation prend de multiples formes selon le stade où elle intervient dans le cycle de programmation (évaluation ex ante, mi-parcours, ex post), selon le niveau de décision concerné (évaluation de politique, de programme, de projet), et selon son caractère plus ou moins étendu (évaluation d'ensemble, évaluation thématique, évaluation plus approfondie de certains aspects d'une politique ou d'un programme). Loin d'être indépendantes les unes des autres, ces différentes formes d'évaluation doivent s'articuler les unes aux autres. Il convient donc de replacer les initiatives d'évaluation dans un cadre plus général, élargi à plusieurs périodes de programmation et à plusieurs niveaux de responsabilité.

En particulier, la succession des périodes de programmation des Fonds structurels européens impose un calendrier qui crée des chevauchements entre les évaluations ex ante, à mi-parcours et ex post. Il en résulte une nécessité de penser la complémentarité entre les travaux pour que chaque évaluation se nourrisse des précédentes, alimente les suivantes et évite les redondances.

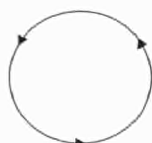
De même, l'emboîtement des niveaux de responsabilité (politique / programme / projet) suppose que les évaluations réalisées à un niveau produisent des conclusions utiles aux autres niveaux. En particulier, les évaluations des programmes prennent en compte les objectifs et les priorités de la politique dans laquelle ils prennent place.

■ Relier les différentes étapes de l'évaluation

Pour bien anticiper l'utilité future d'une évaluation, il faut la replacer dans le cycle de vie du programme considéré comme représenté dans le schéma ci-dessous.

Encadré 6 - L'évaluation dans le cycle de vie du programme

Conception du programme
Evaluation ex ante
Etude de faisabilité



Achèvement du programme
Evaluation ex post
Bilan d'activité

Mise en œuvre du programme
Suivi des réalisations
Evaluation à mi-parcours

● **L'évaluation ex ante**

◀ L'évaluation ex ante intervient en début de cycle alors que le programme n'a pas encore été adopté.

Cette forme d'évaluation contribue à faire en sorte que le programme définitif soit aussi pertinent et cohérent que possible. Ses conclusions sont destinées à être intégrées au programme au moment de la décision.

L'évaluation ex ante porte notamment sur l'analyse des forces, des faiblesses et des potentialités de l'Etat membre, de la région ou du secteur concerné. Elle apporte aux autorités responsables un jugement anticipé sur le fait que les enjeux de développement ont été correctement diagnostiqués, que la stratégie et les objectifs proposés sont pertinents, qu'il n'y a pas d'incohérence avec les politiques et orientations communautaires, que les impacts attendus sont réalistes, etc. Elle apporte par ailleurs les bases nécessaires au suivi et aux futures évaluations en veillant à l'existence d'objectifs explicites et, dans la mesure du possible, quantifiés. Elle aide à préciser les critères de sélection des projets à retenir et à s'assurer du respect des priorités communautaires. Elle contribue, enfin, à assurer la transparence des décisions en permettant d'expliquer clairement les choix opérés et leurs conséquences attendues.

Les évaluations ex ante sont réalisées à un moment où les autorités publiques multiplient les discussions et les négociations à propos du futur programme. Elles sont donc soumises à des contraintes fortes : pression sur les délais, formalisation floue de la proposition de programme à évaluer, modifications de cette proposition en cours de travail, exigences de confidentialité, etc. L'équipe d'évaluation doit par conséquent être capable d'intervenir avec souplesse et rapidité, tout en sachant mettre en œuvre des techniques d'analyse des besoins et de simulation des effets socio-économiques.

● **L'évaluation à mi-parcours**

▶ L'évaluation à mi-parcours intervient au deuxième stade du cycle de programmation, pendant la période de mise en œuvre des interventions.

Au vu de ses conclusions, des ajustements sont apportés, si nécessaire, en milieu de période. Cette évaluation porte un regard critique sur les premières réalisations et sur les premiers résultats des interventions. Elle apprécie également la bonne gestion financière, ainsi que la qualité du suivi

et de la mise en œuvre. Elle montre la traduction opérationnelle des intentions de départ et, le cas échéant, elle constate les modifications apportées de facto aux objectifs initiaux. Par comparaison avec la situation de départ, elle montre l'évolution du contexte économique et social général et juge si les objectifs restent pertinents. Elle examine si l'évolution des politiques et priorités communautaires pose des problèmes de cohérence. Elle aide à préparer des ajustements et des reprogrammations et à les argumenter de façon transparente.

L'évaluation à mi-parcours s'appuie fortement sur les informations issues du système de suivi, mais également sur l'évaluation ex ante, les informations sur le contexte et son évolution. Ce sont en règle générale des exercices relativement courts et exhaustifs qui s'intéressent surtout aux résultats du programme évalué sans chercher à faire une analyse approfondie de ses impacts. En effet, à mi-parcours d'un programme, les impacts n'ont pas souvent eu le temps de se révéler. Il est toutefois possible et souhaitable de se référer à des évaluations approfondies ou thématiques portant sur les programmes antérieurs lorsque ces analyses existent. L'évaluation à mi-parcours a un caractère « endo-formatif », c'est-à-dire qu'elle alimente une rétroaction directe sur le programme dont elle contribue à améliorer la gestion.

● L'évaluation ex post

► L'évaluation ex post récapitule et juge l'ensemble du programme, et notamment ses impacts. Elle vise à rendre compte de l'utilisation des ressources, de l'efficacité et de l'efficience des interventions et de l'obtention des effets attendus et d'autres inattendus. Elle porte sur les facteurs de réussite ou d'échec, sur la durabilité des résultats et impacts. Elle cherche à tirer des enseignements généralisables à d'autres programmes ou à d'autres régions.

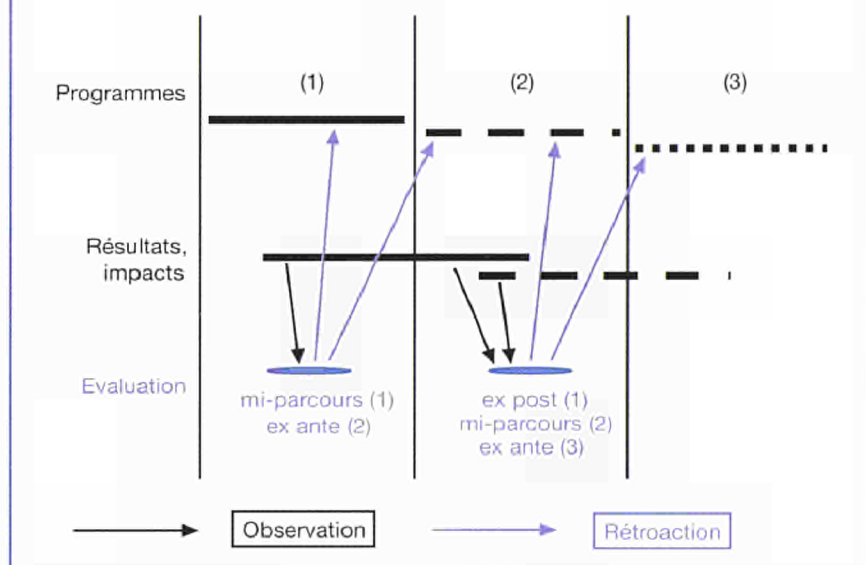
Dans l'idéal, il faudrait disposer de cette évaluation au moment de préparer la prochaine programmation, c'est à dire un an au moins avant la date d'achèvement du programme. Or, pour que les impacts aient eu le temps de se matérialiser, l'évaluation ex post doit être conduite deux à trois ans après la fin de la période de programmation. A défaut, de pouvoir attendre ce délai, il est souvent demandé de réaliser une mise à jour de l'évaluation à mi-parcours, peu de temps avant l'achèvement du cycle de programmation, en liaison avec l'évaluation ex ante du cycle suivant.

L'analyse des impacts est toujours un exercice de grande envergure s'il est conduit systématiquement. Les évaluations ex post sont donc susceptibles d'impliquer des enquêtes de terrain et de se dérouler sur des périodes longues allant de douze à dix-huit mois.

● Les cycles de programmation successifs

L'enchaînement des trois étapes d'évaluation au cours de cycles successifs crée des superpositions qu'il s'agit d'organiser au mieux pour ne pas dupliquer les travaux. Le principe de base revient à adjoindre aux travaux d'évaluation, en cours d'un programme, l'exploitation des conclusions d'évaluations conduites sur le programme précédent. La relative continuité des actions programmées d'une période à l'autre permet, en effet, de tirer des enseignements d'un passé proche sur la pertinence des nouvelles mesures proposées. Le schéma de l'encadré ci-dessous montre que les interactions sont possibles entre les travaux d'évaluation qui interviennent aux différentes phases de plusieurs programmes successifs.

Encadré 7 - Cycles de programmation et évaluations



◀ Ainsi, l'évaluation ex ante qui prépare l'adoption du programme à venir devra mettre à profit les résultats de travaux antérieurs, à savoir :

- L'évaluation à mi-parcours de la période qui s'achève. Cette évaluation aura dégagé les leçons des premières années d'activité et les reprogrammations qui s'en sont suivies. Elle aura éventuellement été complétée par un bilan des réalisations et des résultats du programme en cours, réalisé sur la base des informations du système de suivi.
- L'évaluation ex post de la période antérieure à la période actuelle, éventuellement complétée par des thématiques et des analyses approfondies. Ces évaluations auront permis d'observer et d'analyser les impacts d'interventions déjà anciennes, mais souvent proches des interventions projetées, et intervenant dans un contexte en partie similaire.

► Le rôle de l'évaluation à mi-parcours étant de vérifier si les objectifs sont encore pertinents et sont en train d'être atteints, il y aura lieu de faire principalement référence aux données du système de suivi, mais aussi :

- A l'évaluation ex ante, et notamment au diagnostic réalisé en fonction du contexte socio-économique prévalant avant le début du programme et qu'il y a lieu de mettre à jour.
- A l'évaluation ex post du programme précédent dont les conclusions concernant les mêmes domaines d'interventions pourront servir de références.

► L'évaluation ex post se fonde à la fois sur les données de gestion et de suivi et sur des enquêtes de terrain qui serviront à observer et à analyser les impacts réels et durables des interventions. Elle fait référence à l'évaluation ex ante dans la mesure où elle doit rendre compte de l'atteinte des objectifs. Elle fait référence à l'évaluation à mi-parcours, notamment pour identifier les succès et les échecs qui avaient déjà pu être décelés à ce stade.

Dans la mesure où l'évaluation doit tirer les enseignements de l'expérience des programmes précédents au profit d'une amélioration des programmes suivants, une solution intéressante est d'établir un programme pluriannuel d'évaluation. Il s'agit d'identifier les différentes évaluations possibles et de caler leur date et leur contenu en fonction des calendriers politiques et des échéances décisionnelles des « clients » de l'évaluation aux différents échelons.

■ Mener l'évaluation à différents niveaux

Depuis la réforme des Fonds structurels en 1988, la Commission a privilégié une programmation intégrée des interventions communautaires. Un programme est composé de nombreux projets entrant dans le cadre

d'objectifs spécifiques qui sont eux-mêmes orientés vers l'objectif global du programme.

L'évaluation de programme n'est pas la somme, ni même la synthèse des évaluations des projets qui le composent. Certaines questions ne se posent pas au niveau d'un projet individuel mais intéressent l'évaluation du programme. Cela est par exemple le cas lorsque l'on cherche à évaluer les effets de synergie. Ce point est traité plus loin (voir page 136) ainsi que dans le Volume n°4 de la Collection MEANS).

● Evaluation des projets au stade de leur sélection

Les autorités nationales et régionales ont la responsabilité de la sélection des projets et ceci s'accompagne fréquemment d'une évaluation *ex ante*, également appelée appréciation. L'exemple qui est décrit ci-après est celui d'un programme de développement régional au Royaume-Uni, pays où l'appréciation des projets est réalisée de façon particulièrement soignée.

Exemple : sélection des projets du programme des Highlands and Islands.

Dans le cadre du programme considéré, deux cycles d'appel à projets ont lieu chaque année. Toutes les demandes de financement sont d'abord examinées en fonction de leur éligibilité. Les projets considérés comme éligibles font l'objet d'une notation selon huit critères généraux, communs à l'ensemble du programme (liens avec les objectifs, efficacité probable, respect de l'environnement, additionnalité, innovation, liens avec d'autres projets, intervention du secteur privé et soutien des partenaires locaux) et selon plusieurs critères spécifiques variant d'une mesure à l'autre. Les critères étant pondérés, il est possible d'attribuer une note globale à chaque projet. Après confrontation des demandes de financement et des ressources disponibles, une barre d'admission est établie et une liste de projets à financer est recommandée. La décision finale revient aux responsables du programme.

Cet exemple montre que l'évaluation des projets doit être réalisée en liaison avec les objectifs du programme. Le lien avec les objectifs intervient ici au premier rang des critères de sélection pris en compte dans le système de notation. Inversement, lors de l'évaluation d'ensemble d'un programme, la démarche suit souvent un cheminement du projet au programme, c'est à dire du bas vers le haut (*bottom-up*). Les études sont souvent, dans un premier temps, centrées sur des projets ou des mesures spécifiques, donc de type micro, et étendues progressivement à l'ensemble du programme pour évaluer les hypothèses de causalité qui

sont, par nature, toujours plus macro. L'évaluation des projets nourrit alors l'évaluation de programme.

Exemple : évaluation d'un projet dans son contexte régional.

Une équipe d'évaluation a réalisé l'évaluation ex ante d'une aide publique attribuée pour l'implantation d'une usine de construction automobile en Italie du sud. La question principale portait sur les retombées économiques et sociales de cette usine pour l'activité économique locale, et notamment pour la réduction du chômage. En examinant le projet de façon isolée, l'équipe d'évaluation a montré que les retombées étaient faibles. En fait, l'usine restait fortement dépendante de ressources externes à la région en termes de main œuvre et de fournisseurs. En ne considérant que le niveau microéconomique, l'impact du projet était donc limité.

En examinant, au contraire, le projet dans son contexte régional, l'équipe d'évaluation a montré que le projet était susceptible de générer des activités nouvelles, de développer des productions locales de biens et services, et finalement de produire des effets significatifs à long terme, à condition de prendre en compte des mécanismes de niveau macroéconomique.

Les limites de la démarche qui va du micro au macro réside dans le fait qu'elle entraîne l'équipe d'évaluation à faire des hypothèses de plus en plus fragiles dans la mesure où les effets abordés sont de plus en plus lointains et complexes. Malgré tout, l'avantage de cette articulation des évaluations de projets et de programme est de fournir une explication détaillée de l'enchaînement des impacts à court et moyen termes.

● **Evaluation des grands projets**

Un cas particulier est fait pour l'évaluation des grands projets d'infrastructure qui constituent des instruments importants de la mise en œuvre de la politique régionale de l'Union européenne. Les grands projets ont été définis comme « ceux dont le coût total [...] excède, en général, 25 millions d'euros pour les investissements en infrastructure et 15 millions d'euros pour les investissements productifs »². L'importance des financements mis en jeu justifie que leur appréciation ex ante soit menée de façon approfondie et autonome, en comparant leur coût et leurs avantages selon une technique décrite dans le Volume 3 de la Collection. La responsabilité de ces évaluations incombe complètement aux Etats membres³.

² Article 16 du Règlement du Conseil 2082/93

³ Règlement 2082/93.

Pour les investissements en infrastructures l'analyse des avantages socio-économiques du projet inclut le taux prévisible d'utilisation, l'impact potentiel sur le développement ou la reconversion de la région concernée et l'estimation des conséquences de l'intervention communautaire sur la réalisation du projet. Pour les investissements productifs, on analysera plus particulièrement les perspectives du marché dans le secteur concerné, les effets sur l'emploi et l'analyse de rentabilité économique et financière prévisionnelle du projet.

Dans l'instruction des demandes de concours individuels, la Commission prend en compte les résultats des appréciations et des évaluations effectuées. Si l'appréciation présentée par le demandeur est considérée comme insuffisante et peu convaincante, la Commission peut réclamer au demandeur qu'il développe l'analyse, ou bien réaliser elle-même une appréciation du projet en sollicitant, si nécessaire, l'avis d'experts techniques indépendants. Dans tous les cas, la décision de la Commission sera le résultat d'un dialogue et d'un travail commun avec le demandeur, afin de parvenir aux meilleurs résultats d'investissements.

● **Evaluation de programme**

Les résultats des évaluations de programme sont systématiquement repris dans des synthèses réalisées à l'échelon national et à l'échelon européen. A ces échelons, elles servent à alimenter la réflexion et à aider aux prises de décisions relatives à la politique de cohésion économique et sociale. L'encadré suivant liste une série de « clients » directs et indirects des évaluations de programme à l'échelon européen.

Encadré 8 - Les « clients » des évaluations à l'échelon européen

Les évaluations de programme sont directement ou indirectement utiles :

- aux exécutifs européens pour alimenter leur réflexion stratégique sur la politique de cohésion et pour préparer leurs propositions au Conseil,
- aux unités opérationnelles de la Commission, dans le cadre de la préparation des décisions relatives au financement des programmes et à leur contenu,
- aux représentants de la Commission, lorsqu'ils expriment le point de vue communautaire lors des décisions de réorientation au cours de la mise en œuvre du programme,
- et enfin aux services chargés de la préparation des rapports annuels sur la mise en œuvre des Fonds structurels et du rapport triennal sur la cohésion économique et sociale.

Une forme d'évaluation particulièrement intéressante à l'échelon européen est celle qui analyse certains impacts considérés comme prioritaires à cet échelon. Ainsi l'évaluation peut aborder des questions transversales telles que l'impact sur l'emploi, la compétitivité des PME, l'égalité des chances entre les femmes et les hommes, l'environnement, etc. Ces priorités communautaires sont souvent rappelées dans les documents de programmation sans pour autant que l'on puisse parler d'une intégration véritable (mainstreaming). Chercher à évaluer les impacts en relation avec ces priorités revient à concevoir ce que l'on appelle une évaluation thématique (voir le Volume 5 de la Collection).

Une évaluation thématique proprement dite porte sur plusieurs programmes et peut être réalisée à l'échelon national ou européen. Sa particularité sera de traiter un thème spécifique, en général central pour le commanditaire. Par exemple, la Commission européenne a réalisé une évaluation sur le thème de l'égalité des chances, dans une dizaine de programmes répartis dans des Etats différents. Les évaluations thématiques permettent de faire un rapprochement entre différents programmes et d'avoir une meilleure compréhension des processus de production d'impact dans un domaine particulier. De grandes évaluations transnationales ont ainsi été lancées sur l'Environnement, la Recherche-Développement, le soutien aux Petites et Moyennes Entreprises.

■ Sélectionner les différents niveaux d'analyse

L'expérience d'évaluation des programmes des Fonds structurels montre une grande diversité dans la manière de traiter les questions évaluatives. Assez souvent, il est demandé à l'équipe d'évaluation de traiter un ensemble de questions, dans le cadre d'un exercice limité à quelques mois. D'autres évaluations consacrent au contraire une année entière à l'analyse approfondie d'un seul aspect. Le choix d'approfondir une question sera guidé par le besoin de mieux comprendre une difficulté rencontrée, de mieux montrer le caractère exemplaire d'une réussite, de mieux connaître l'efficacité d'une mesure, de mieux comprendre les conditions d'atteinte d'un objectif, etc. Dans l'exemple présenté en page suivante, le groupe de pilotage a envisagé d'approfondir plusieurs questions prioritaires.

C'est la combinaison de plusieurs critères qui va conduire à la sélection des questions à évaluer en profondeur. Il pourra s'agir :

- des mesures les plus coûteuses (en supposant que les enjeux décisionnels augmentent avec le budget) ;

- des mesures les plus innovantes (en considérant que ces parties du programme sont des expérimentations qui ont besoin d'être validées par une évaluation) ;
- des objectifs les plus stratégiques (en supposant que ces parties du programme soient le plus à même de profiter d'un débat informé) ;
- des objectifs les plus controversés (en supposant que ces parties du programme soient le plus à même de profiter d'une clarification des critères de jugement) ;
- des priorités qui intéressent particulièrement un échelon de décision.

Exemple : sept questions à approfondir dans le cadre de l'évaluation d'un programme

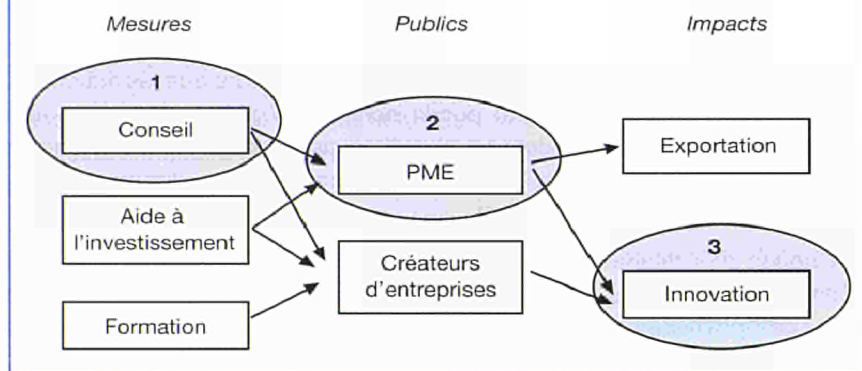
En 1997, dans le cadre de l'évaluation d'un programme de reconversion industrielle en Franche-Comté au nord-est de la France, le groupe de pilotage de l'évaluation a consacré ses deux premières réunions à identifier une série de questions et à les hiérarchiser.

Sept champs de questionnement ont ainsi été identifiés qui ont fait l'objet de discussions en ateliers afin de formuler, pour chacun d'eux, une question évaluative réaliste et utile. Dans l'ordre des priorités établies par le groupe de pilotage, ces questions sont les suivantes : (1) Quelles catégories de projets ont le meilleur impact sur l'attrait du territoire ? (2) Les aides aux entreprises ont-elles permis de créer ou de maintenir des emplois fortement enracinés dans la zone ? (3) Les services aux entreprises ont-ils suffisamment contribué à la diversification de l'activité économique ? (4) Les formations servent-elles à élever les qualifications ou plutôt à les rendre plus pertinentes ? De ces deux effets attendus, lequel est le plus nécessaire ? (5) L'impact probable des projets d'enseignement supérieur et de recherche est-il suffisant ? (6) Quelle est l'efficacité des différentes catégories de projets en termes de création et de maintien d'emploi ? (7) Quelles catégories de projets assurent ou améliorent l'égalité des chances entre les femmes et les hommes ?

L'approfondissement des sept questions aurait représenté un coût trop élevé. Il a donc été décidé de limiter le nombre de questions à approfondir à trois. Certaines questions ont été écartées dans le souci de ne pas créer de doublon avec d'autres évaluations en cours dans la région. Finalement les questions (1), (4) et (5) ont été jugées prioritaires.

Comme le montre l'Encadré 9, les analyses approfondies peuvent prendre différentes formes qui dépendent de l'objet analysé.

Encadré 9 – Trois situations d'analyse approfondie



La première forme d'analyse approfondie (cercle 1 dans l'Encadré) concerne des impacts d'une mesure, ou d'un grand projet. En fonction de la nature du public concerné, les impacts recherchés seront par exemple, les emplois créés ou maintenus, l'investissement induit, la consolidation des activités, la diversification des activités, etc. Les destinataires de l'intervention sont connus et peuvent être interrogés. Les données primaires ainsi obtenues fournissent des indications relatives aux effets bruts. L'analyse de causalité peut également être conduite à l'aide des techniques proposées dans le Volume 3 (voir page 157) : recherche des facteurs concurrents, estimation des effets d'aubaine, de déplacement ou d'éviction, constitution d'un groupe de comparaison, construction d'une situation contrefactuelle. Dans cette situation, l'évaluation aboutit à des conclusions sur le succès ou l'échec d'une intervention donnée. Ces conclusions seront si possible, généralisables à d'autres interventions similaires (validité externe).

La deuxième forme d'analyse approfondie (cercle 2 dans l'Encadré) se concentre sur un public particulier, comme par exemple les chômeurs de longue durée, les créateurs d'entreprise, les petites exploitations agricoles, etc. L'évaluation se place du point de vue du public concerné, analyse ses besoins, son attitude par rapport au programme, ses critères de jugement. Dans cette situation, l'évaluation aboutit à des conclusions sur l'impact comparé des différentes mesures qui touchent directement ce public. Elle peut également montrer comment ces différentes mesures se renforcent ou se contrarient réciproquement (effets de synergie).

La troisième forme d'analyse approfondie (cercle 3 dans l'Encadré) concerne un impact ou une catégorie d'impacts. La question s'applique alors à toutes les mesures du programme et porte sur leur contribution respective à la production de cet impact. La particularité de ce type de situation est que le public concerné est homogène, alors que les actions qui exercent une influence sur ce public sont hétérogènes et nombreuses. Dans cette situation, l'évaluation aboutit à des conclusions sur l'impact comparé des différentes mesures qui contribuent directement ou indirectement à produire l'impact.

La notion d'évaluation thématique évoque celle d'évaluation approfondie. Certaines évaluations thématiques examinent comment un même impact (ex : compétitivité des PME) est obtenu par plusieurs sortes d'interventions (voir Encadré 9 - situation 3), ou comment les interventions financées dans un domaine précis comme la Recherche & Développement produisent plusieurs sortes d'impacts (voir situation 1).

Les évaluations thématiques peuvent être réalisées au sein d'un programme (le terme est alors équivalent à celui d'analyse approfondie) ou de façon transversale à plusieurs programmes. Ces évaluations sont décrites dans le Volume 5 de la Collection MEANS.

Les évaluations thématiques sont particulièrement utiles pour évaluer les programmes du point de vue des priorités communautaires. Elles reposent sur le principe du « *mainstreaming* » qui repose sur la prise en compte systématique d'une priorité ou de besoins spécifiques dans l'ensemble des politiques communautaires, et en particulier dans le programme évalué. Le concept est particulièrement employé pour la prise en compte des besoins des femmes et des hommes dans l'ensemble des politiques afin de promouvoir un objectif d'égalité des chances qui constitue une priorité. Cet objectif peut être atteint si l'on tient compte de l'effet de l'intervention sur la situation spécifique des femmes et des hommes dès la conception du programme, puis au cours de la mise en œuvre, du suivi et de l'évaluation, et jusque dans les systèmes d'indicateurs. Le *mainstreaming* nécessite le développement de mesures visant à adapter l'organisation de la société pour garantir une distribution plus équitable des rôles entre les femmes et les hommes

■ Différencier le contrôle, le suivi et l'évaluation

Il faut attendre de l'évaluation ce qu'elle peut réellement apporter et ne pas chercher à lui faire jouer des rôles qui ne sont pas les siens, tels que le contrôle et le suivi.

Dans le cadre des politiques structurelles, la Commission se préoccupe de satisfaire simultanément les exigences du Conseil et du Parlement en matière de contrôle, de suivi et d'évaluation. Les trois exercices sont proches en tant que jugements portés sur l'activité publique et peuvent s'enrichir mutuellement. Toutefois, ils doivent être clairement distingués pour éviter toute ambiguïté. Ces pratiques requièrent, en effet, des qualités professionnelles, des temporalités et des modes d'organisation différents.

Une première clarification concerne le point de vue à partir duquel l'action publique est jugée. Le contrôle vérifie la légalité et la régularité de la mise en œuvre des ressources. Le suivi vérifie la bonne gestion des interventions et produit une analyse régulière sur l'état d'avancement des réalisations. L'évaluation juge de la mise en œuvre du programme sur la base des réalisations, des résultats et des impacts qu'il a produits dans la société.

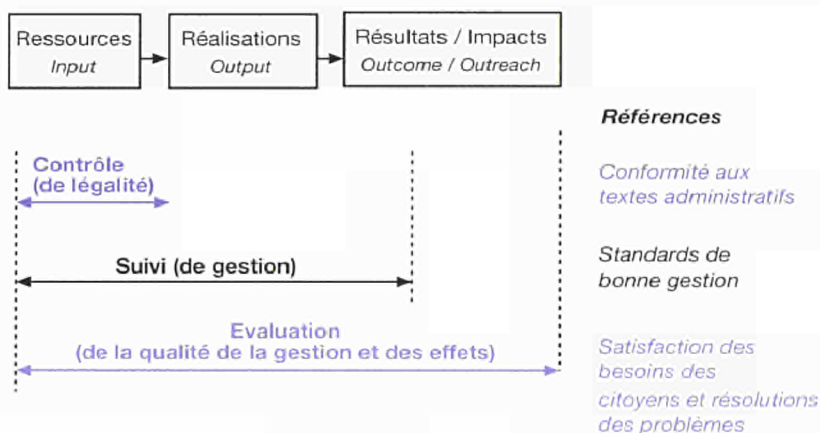
Un deuxième éclaircissement concerne les critères de jugement. Le contrôle juge en fonction de critères connus et clarifiés à l'avance (enveloppes budgétaires, textes réglementaires, normes professionnelles). Le suivi juge en fonction d'objectifs opérationnels à réaliser. Au contraire, l'évaluation doit le plus souvent commencer par choisir ses critères de jugement. Ceux-ci sont formulés à partir des objectifs de l'action publique évaluée. Or l'expérience nous montre, et la littérature de l'évaluation le confirme : il est encore rare que les objectifs d'une action publique soient clairs, quantifiés, vérifiables et immédiatement utilisables comme critères d'évaluation.

Ces différences fondamentales qui distinguent contrôle, suivi et évaluation sont présentées dans le schéma en page suivante.

Ainsi, les qualités professionnelles et les compétences requises par les différents exercices ne sont pas les mêmes. Le contrôleur a une connaissance approfondie des critères préétablis (connaissances juridiques ou comptables par exemple). Son comportement inspire le respect et l'autorité et il n'hésite pas à utiliser la contrainte et à stigmatiser publiquement les erreurs et les dysfonctionnements pour obtenir un effet de dissuasion. Le responsable du suivi doit avoir pour sa part des

compétences reconnues en termes d'organisation, de gestion et de mise en place des dispositifs d'information. Il supervise et assiste les opérateurs dans la mesure où il leur fournit des outils et des procédures qui les aident dans leurs pratiques professionnelles et qui de plus serviront à une analyse critique des conditions de mise en œuvre. Enfin, le praticien de l'évaluation ne pouvant que difficilement s'appuyer sur des critères unanimement reconnus, a besoin de la collaboration des gestionnaires publics pour collecter une information pertinente, pour savoir l'interpréter et pour formuler un jugement légitime. Il ne peut pas être efficace si ses interlocuteurs ne sont pas de bonne volonté et de bonne foi, ce qui implique souvent de respecter une certaine confidentialité.

Encadré 10 - Articuler contrôle, suivi et évaluation



Si les trois exercices présentent des différences intrinsèques, ils se servent mutuellement grâce à leur complémentarité. Par exemple, le contrôle et le suivi peuvent repérer des succès et des échecs apparents, que l'évaluation va chercher à confirmer par une analyse approfondie. A l'inverse, une évaluation peut montrer que l'application d'une norme juridique ou professionnelle constitue un facteur particulier d'efficacité ou d'inefficacité. Entre les exercices de suivi et d'évaluation il existe une interdépendance en ce sens que les indicateurs du suivi sont utiles à l'évaluation.

Toutefois, les risques de conflits entre les trois logiques sont importants. En particulier, une évaluation ne doit pas être réalisée avec les méthodes du

contrôle sous peine de créer des oppositions qui contrarient la collecte d'information et l'utilisation des conclusions. Dans un autre ordre d'idées, aucune évaluation ne peut remplacer un contrôle ou un suivi défaillant. Par exemple, en cas de soupçon sur la régularité ou le professionnalisme de la mise en œuvre, il est nécessaire de procéder à une inspection ou à un audit afin de prendre les décisions qui s'imposent et d'attendre que la sérénité soit revenue pour investir dans une évaluation.



Décider le lancement de l'évaluation

La plupart des évaluations entrant dans le cadre de la politique européenne de cohésion économique et sociale ont un caractère obligatoire. Cet état de fait ne devrait pas conduire à sous-estimer le rôle que joue la décision de lancement d'une évaluation. Celle-ci est l'occasion de cadrer le travail à entreprendre en définissant précisément son champ (ce qui va être évalué), son calendrier et surtout ses motifs réels, au-delà du simple respect des obligations réglementaires.

En formulant une décision claire et forte, les autorités publiques commanditaires donnent à l'évaluation toute sa légitimité politique. Elles désignent les personnes qui participeront au processus d'évaluation en répartissant les rôles et en fixant les règles du jeu.

Lorsque l'évaluation est obligatoire, on peut craindre que les autorités commanditaires soient tentées par une solution de facilité qui consiste à calquer leurs décisions sur des modèles préétablis sans se soucier suffisamment de ce qu'ils peuvent réellement attendre de l'évaluation dans le contexte de leur programme et dans les limites du calendrier politique. Il en résulte un risque d'uniformisation et de bureaucratisation de l'évaluation. Plusieurs auteurs de référence dans le domaine de l'évaluation affirment, au contraire, que chaque évaluation doit être considérée comme unique (Patton, 1986) parce qu'elle répond à des attentes qui changent en fonction du moment et du contexte, mais aussi parce que l'expérience a montré qu'il n'existe pas de méthode universelle d'évaluation.

■ Délimiter le champ

La décision d'évaluer est l'occasion de définir les limites du programme en termes de dimensions institutionnelles, temporelles, sectorielles et géographiques. C'est ce que l'on appelle le champ de l'évaluation ou « évaluanda ». Délimiter le champ d'une évaluation revient à se poser la question : « Que veut-on évaluer ? ». Le champ d'investigation doit être circonscrit sur quatre plans au moins : institutionnel (niveau européen, national ou local), temporel (période prise en considération), sectoriel (social, industriel, environnement, rural, etc.) et géographique (telle partie du territoire européen, telle région, telle ville, tel parc naturel, etc.).

Dans le cadre de cet ouvrage, le champ est en principe délimité par les financements du programme à évaluer, par le territoire concerné et par la

période de programmation. Il est cependant utile d'avoir une réflexion préalable sur les points suivants :

- Veut-on évaluer les financements du programme uniquement ou également certains autres financements nationaux, régionaux ou locaux qui sont plus ou moins directement liés au programme ?
- Veut-on évaluer les interventions dans la zone éligible uniquement ou étendre d'observation à certaines zones voisines qui rencontrent des problèmes de développement similaires ?
- Veut-on évaluer les financements attribués uniquement dans le cycle de programmation concerné ou également certains financements des cycles précédents ?

Il faut insister ici sur l'importance d'une délimitation relativement stricte du champ. L'expérience a en effet montré qu'au cours du processus d'évaluation les participants ont tendance à vouloir examiner tout ce qui touche de près ou de loin au programme. Pour aboutir à des conclusions, l'évaluation doit être réduite à l'examen du programme et de ses interdépendances les plus essentielles avec d'autres politiques publiques.

◀ Le risque est particulièrement sensible dans le cas des évaluations ex ante qui peuvent dériver vers un exercice d'analyse prospective ou de réflexion spéculative qui s'éloigne de l'objet de l'évaluation. Il est recommandé de bien limiter le champ de l'évaluation à la proposition de programme.

■ Annoncer les motifs et l'utilisation prévue

Au moment du lancement de l'évaluation, il est fortement recommandé de préciser les motifs qui sont à l'origine de la décision d'évaluer et de dresser la liste des utilisateurs potentiels. Même si l'évaluation est obligatoire, cet exercice est une façon de s'assurer contre la production d'un travail inutile. Il est également important de préciser jusqu'où on souhaite voir l'équipe d'évaluation s'impliquer. Est-ce que les attentes se limitent à de simples conclusions ou est-ce que l'on souhaite obtenir des recommandations à caractère opérationnel ou stratégique ? Les attentes sont évidemment différentes selon l'étape à laquelle l'évaluation prend place (ex ante, intermédiaire ou ex post).

Il faut souligner ici l'importance, pour obtenir une évaluation utile, d'énoncer, avec le plus de précision possible, les décisions susceptibles

d'être prises et qui peuvent être éclairées par l'évaluation. On constate fréquemment que les commanditaires n'osent pas, par peur de trop influencer l'équipe d'évaluation, exprimer par avance les changements qui leur paraissent souhaitables d'apporter au programme ou leurs doutes par rapport à l'efficacité de telle ou telle action. L'intention est louable : ne rien dévoiler par avance pour voir si l'équipe d'évaluation arrive aux mêmes conclusions que les siennes ! L'expérience montre au contraire que l'évaluation a peu de chance de documenter les décisions que l'on envisage de prendre si celles-ci ne sont pas connues au préalable de ceux qui vont aller sur le terrain collecter et analyser les données. La réalité socio-économique est en effet très complexe et l'équipe d'évaluation se trouve confrontée à un très grand nombre de constats et de pistes d'amélioration possibles. Elle risque donc de s'attacher à vérifier des hypothèses certainement intéressantes dans l'absolu mais peu réalistes aux yeux des responsables politiques ou des gestionnaires compte tenu des marges de manœuvre existantes ou des urgences.

■ Répartir les rôles dans le dispositif d'évaluation

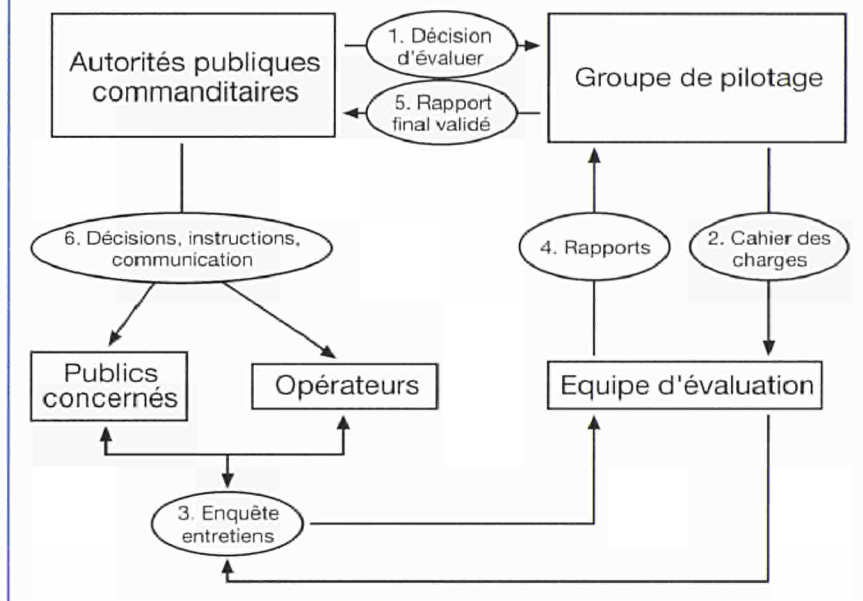
On entend par « dispositif d'évaluation » les rôles et responsabilités au sein du processus d'évaluation d'une part et, d'autre part, les relations entre l'évaluation, le suivi et la gestion du programme. Des règles du jeu claires et équitables contribuent à des travaux de qualité et évitent les blocages. Le choix principal porte sur le rôle de pilotage de l'évaluation, rôle qui comprend notamment la rédaction du cahier des charges, la sélection de l'équipe d'évaluation et la validation de son rapport.

Par le passé, le pilotage de l'évaluation a été souvent confié à un fonctionnaire isolé ou au contraire, exercé collectivement par les cofinanceurs du programme dans le cadre de réunions formelles, élargies à de nombreux partenaires et peu fréquentes (Comités de suivi). En pratique, le dispositif fonctionne bien quand la responsabilité du pilotage est assurée, de façon relativement informelle, par un groupe restreint composé des fonctionnaires des administrations les plus directement concernées par le programme, auquel on adjoint parfois quelques experts. La périodicité des réunions doit être importante (une par mois à certains moments) et les procédures de travail doivent être souples et peu formelles.

De façon schématique (voir schéma ci-après) on peut réduire le dispositif d'évaluation à la répartition des rôles entre les cinq entités clés que sont les

commanditaires, le groupe de pilotage, l'équipe d'évaluation et enfin les opérateurs et les publics concernés. Dans certaines situations il convient de rajouter d'autres entités. En particulier on a parfois, avec succès, organisé le pilotage sur la base d'un comité global coordonnant plusieurs comités sectoriels (ou des comités créés à une échelle géographique plus petite). Les responsabilités de ces entités sont décrites par leur rôle respectif au cours des six étapes clés qui constituent l'évaluation.

Encadré 11 – Le dispositif d'évaluation et ses six étapes



- Etape 1 : Les autorités publiques, commanditaires et futurs utilisateurs de l'évaluation, formulent leurs instructions à un groupe de pilotage. Cette étape se matérialise a minima par un paragraphe dans le document de programmation ; elle pourrait à l'avenir faire l'objet d'un mandat en bonne et due forme ainsi qu'il est recommandé dans le chapitre de conclusion de ce Volume.
- Etape 2 : Le cahier des charges est conçu et rédigé par le groupe de pilotage sur la base du mandat reçu. Ce cahier des charges sert le plus souvent de document d'appel d'offre et permet la sélection de l'équipe d'évaluation. Le groupe de pilotage assurera autant d'interactions que

nécessaires avec l'équipe d'évaluation afin de s'assurer que celle-ci ne dévie pas des lignes directrices inscrites au cahier des charges et travaille dans le sens des questions posées.

- Etape 3 : L'équipe d'évaluation réalise la collecte d'informations secondaires et primaires auprès des opérateurs et des publics concernés puis en fait l'analyse. C'est l'étape la plus longue.
- Etape 4 : Un rapport que l'on peut qualifier de « technique » est rédigé par l'équipe d'évaluation. Il inclut des conclusions et le cas échéant des recommandations provisoires à l'intention du groupe de pilotage.
- Etape 5 : Les conclusions sont validées en tout ou partie par le groupe de pilotage qui en outre formule ses recommandations aux autorités publiques dans le rapport d'évaluation qui intègre le rapport technique ou le présente en annexe.
- Etape 6 : Les autorités utilisent l'évaluation en la diffusant auprès du public et/ou en décidant de nouvelles orientations que les administrations seront chargées de mettre en œuvre. Cette étape cruciale est souvent isolée du processus d'évaluation proprement dit car le groupe de pilotage peut être dissous après avoir rendu son rapport et ne peut malheureusement pas assurer un véritable suivi des décisions subséquentes. Notons toutefois que l'on peut se prémunir contre le risque d'absence de suites données en incluant dans le mandat initial un engagement moral des autorités d'inviter si ce n'est le groupe de pilotage au moins son Président à présenter les conclusions et à assister à la réunion au cours de laquelle seront prises les décisions.

◀ Il faut souligner que par le passé les dispositifs d'évaluation ont peu à peu convergé vers le schéma qui précède pour les évaluations à mi-parcours. Par contre tel n'a pas été le cas des évaluations ex ante. Les contraintes de temps et la difficulté de mobiliser les protagonistes par rapport à un programme qui n'existe pas encore peuvent expliquer sans toutefois complètement justifier l'absence de groupe de pilotage.

► Les mêmes lacunes se retrouvent en matière d'évaluation ex post alors même que l'on ne peut se prévaloir de difficultés similaires. Ce type d'évaluation est souvent demandé par une administration sans véritablement apporter tout le soin nécessaire à la clarification des attentes des niveaux décisionnels les plus importants ou au pilotage du processus.

■ Formuler les questions évaluatives

● Liste des questions possibles

L'aspect qui est peut-être le plus fondamental dans la phase de conception d'une évaluation est la formulation des questions évaluatives. Ce sont ces questions qui constituent le cœur de l'exercice d'évaluation.

Parmi les questions possibles, on distingue en général les questions descriptives qui cherchent à observer et à mesurer des changements (que s'est-il passé ?), les questions causales qui cherchent à comprendre et apprécier des liens de cause à effet (comment et dans quelle mesure ce qui s'est passé est attribuable au programme ?), et enfin les questions normatives qui appliquent des critères d'évaluation (les résultats et les impacts sont-ils satisfaisants ?). Dans l'idéal, toute question évaluative devrait englober ces trois dimensions et avoir les qualités suivantes :

La question doit correspondre à une véritable interrogation, à un réel besoin d'information, de compréhension et/ou d'identification de solutions nouvelles. Dans le cas contraire, il s'agit d'une « question alibi », visant par exemple à cautionner une décision déjà prise. Si une question présente un intérêt uniquement en terme de connaissances nouvelles sans débouché immédiat en termes de décisions ou de débat publics, elle relève plus de la recherche scientifique, elle ne trouve pas sa place dans un exercice d'évaluation.

La question concerne un impact, un groupe d'impacts, un résultat ou un besoin, c'est à dire qu'elle concerne au moins en partie des éléments extérieurs au programme, à savoir ses destinataires ou son environnement économique et social. Si une question concerne uniquement la gestion interne des ressources et des réalisations, il est probable qu'elle peut être mieux traitée dans le cadre des exercices de suivi ou de contrôle.

Exemple : Quelques questions évaluatives relatives aux programmes des Fonds structurels européens

Les aides aux centres scientifiques et techniques ont-elles provoqué un début de rattrapage du retard des PME régionales en matière d'innovation ?

Le programme a-t-il créé ou maintenu un nombre suffisant d'emplois au regard de son coût ?

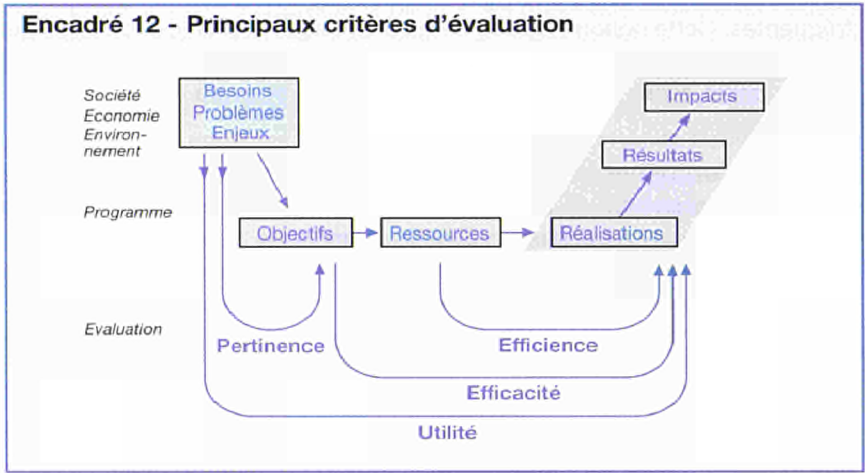
Les réalisations et les résultats du programme correspondent-ils aux besoins et aux attentes spécifiques des femmes ?

Quelles sont les mesures qui ont les impacts les plus positifs (les plus négatifs) sur l'environnement ?

La question comporte un critère de jugement et un seul. Ce dernier élément pourra paraître parfois difficile à respecter, mais l'expérience montre pourtant qu'il s'agit d'un facteur clé pour l'utilité de l'évaluation. En l'absence de critère de jugement annoncé dès le départ, le rapport final parvient rarement à déboucher sur des conclusions.

● Critères d'évaluation

Lorsqu'elles comprennent un critère de jugement, les questions d'évaluation se rangent principalement dans l'une des quatre catégories suivantes : celles qui ont trait à la pertinence du programme, celles qui ont trait à son efficacité, celles qui ont trait à son efficience ou encore celles qui ont trait à son utilité. Ces quatre grandes catégories sont représentées dans le schéma ci-dessous.



◀ ▶ Par « pertinence » on entend, dans le contexte d'une évaluation, l'adéquation entre les objectifs explicites du programme et les problèmes socio-économiques que ce dernier est censé prendre en charge. Dans le cadre d'une évaluation ex ante, les questions de pertinence sont les plus importantes car on s'interroge principalement sur la stratégie à choisir ou sur la justification de celle que l'on propose. Dans le cadre d'une évaluation à mi-parcours, on souhaitera vérifier si le contexte socio-économique a bien évolué comme on l'imaginait a priori et si cette évolution ne remet pas en question tel ou tel objectif initial.

►► Les questions fondamentales des évaluations à mi-parcours et plus encore des évaluations ex post sont relatives à l'efficacité et à l'efficience des interventions mises en œuvre et du programme dans son entier. En ce qui concerne l'efficacité, il s'agit de savoir si les objectifs formulés dans le programme sont en voie d'être atteints, de connaître les réussites et les difficultés rencontrées, l'adéquation des solutions choisies ou encore l'influence de facteurs concurrents exogènes au programme lui-même.

L'efficience dans le contexte de l'évaluation revient à rapprocher les résultats obtenus, ou mieux encore les impacts produits des ressources mobilisées. Autrement dit, les effets obtenus sont-ils à la hauteur des moyens engagés ? Les termes d'efficience et d'efficacité sont, par ailleurs, couramment employés par les gestionnaires dans une acception différente dans la mesure où ils s'attacheront, dans le cadre du suivi, à l'analyse des réalisations et non des résultats ou impacts. Soulignons à cet égard que les questions relatives à la performance d'un programme sont de plus en plus fréquentes. Cette notion englobe celles d'efficacité et d'efficience telles que nous les avons définies à la fois pour le suivi et pour l'évaluation. C'est donc un ensemble assez large de questions que l'on trouvera regroupées sous ce vocable.

Le critère d'utilité juge les impacts obtenus par le programme au regard des besoins de la société et des enjeux économiques. L'utilité est un critère d'évaluation très particulier puisqu'il abandonne toute référence aux objectifs officiels du programme. Il peut être judicieux de formuler une question d'utilité lorsque les objectifs sont mal définis ou lorsque qu'il y a beaucoup d'impacts imprévus. Ce critère doit malgré tout être utilisé avec précaution afin d'éviter que l'équipe d'évaluation ne procède à une sélection trop personnelle des besoins ou enjeux importants. Certains auteurs nord-américains ont prôné cette forme d'évaluation « affranchie des objectifs » (*goal free evaluation*).

De façon concise on peut caractériser simplement les quatre catégories de questions de la façon suivante :

- Pertinence : Dans quelle mesure les objectifs du programme sont-ils justifiés par rapport aux besoins ? Leur raison d'être est-elle encore démontrable ? Correspondent-ils aux priorités locales, nationales et européennes ?
- Efficacité : Dans quelle mesure les objectifs escomptés ont-ils été atteints ? Est-ce que les instruments mis en œuvre ont produit les effets attendus ? Pourrait-on obtenir plus d'effets en utilisant des instruments différents ?

- **Efficience** : Est-ce que les objectifs ont été atteints au meilleur coût ? Pourrait-on obtenir plus d'effets pour le même coût ?
- **Utilité** : Est-ce que les effets attendus et inattendus sont globalement satisfaisants du point de vue des destinataires directs et indirects ?

Ces quatre critères ne sont pas exclusifs. Par exemple, on posera une question du type : Est-ce que les effets obtenus vont perdurer au-delà de la fin du programme, même en cas d'arrêt de tout financement public ? Il s'agira dans ce cas du critère de viabilité (*sustainability*) ou de durabilité des impacts produits. On peut encore avoir recours à d'autres critères, tels que : l'équité, la cohérence, la synergie, la reproductibilité, etc.

● Probabilité de réponse

Une fois les questions évaluatives identifiées, il convient de réfléchir à leur évaluabilité. Il s'agit d'apprécier *a priori* si les questions évaluatives peuvent vraisemblablement trouver des réponses compte tenu des ressources disponibles. Il s'agit concrètement de s'assurer qu'une équipe d'évaluation utilisant des outils d'évaluation appropriés sera capable, dans des délais et pour un coût compatibles avec les contraintes existantes, de trouver tous les éléments nécessaires à la construction de réponses crédibles aux questions posées. Pour savoir s'il est possible d'accéder à un degré de crédibilité acceptable, on peut faire, ou faire faire une étude d'évaluabilité. On y reviendra dans la dernière partie de l'ouvrage (voir page 211) consacrée aux améliorations qui peuvent être apportées aux pratiques existantes.

Il est recommandé de vérifier, même de façon sommaire, pour chaque question évaluative :

- si les concepts sont stabilisés ;
- si l'on saura formuler des hypothèses explicatives ;
- si les données déjà disponibles permettent de répondre à la question sans enquête complémentaire ;
- si l'accès au terrain ne posera pas de difficultés majeures.

Un certain nombre de facteurs peuvent rendre très difficile la réponse à une question. Ainsi, par exemple, si le programme est très nouveau, s'il n'a pas encore pu produire de résultats importants ou si les données disponibles sont totalement inappropriées. Ces raisons peuvent conduire à la décision de ne pas entreprendre l'évaluation, de la repousser à plus tard, ou encore de poser des questions plus réalistes.

■ Sélection des questions prioritaires

Parce que chercher à répondre de façon approfondie à toutes les questions serait trop coûteux en temps et en investigations, il est toujours recommandé d'établir une hiérarchie parmi l'ensemble des questions évaluatives exprimées.

Pour ce faire, il est nécessaire d'apprécier les enjeux associés à chacune de ces questions (connaissance, négociation, décision, communication) afin de pouvoir établir une hiérarchie entre elles. Les probabilités de réponse aux questions et les probabilités d'utilisation de ces réponses doivent guider le choix qui sera d'autant plus « réussi » que les utilisateurs potentiels auront été consultés et se seront appropriés les questions sélectionnées.

Il est bien entendu qu'il vaut mieux obtenir des réponses légèrement imprécises à des questions importantes que des réponses précises à des questions que personne ne se pose et qui risquent *a fortiori* de ne jamais être écoutées.

Encadré 13 - Sélectionner les questions prioritaires

Parties prenantes
Types de décisions
à prendre
Contexte politique

Première liste de
questions évaluatives :

Probabilité de réponse

Probabilité d'utilisation

Questions
prioritaires

Comme le montre l'encadré ci-dessus, on considérera qu'une question est prioritaire si la probabilité de réponse est grande et si la probabilité d'utilisation de la réponse l'est aussi. A cette fin, on se posera les interrogations suivantes :

- Les conclusions seront-elles utilisées ? Par qui ? Pour quoi faire (décider, débattre, rendre compte) ? Quand ?
- Est-il politiquement opportun de mener une telle évaluation à un tel moment ou dans un tel contexte ? Y a-t-il une situation conflictuelle qui risque de compromettre l'exercice ?
- Une étude récente n'a-t-elle pas déjà répondu à la majorité des questions ?
- Les priorités en matière d'évaluation ne sont-elles pas trop instables ?

Si nécessaire, ces points peuvent faire l'objet d'une étude préalable à l'évaluation (voir page 200).



Passer commande de l'évaluation

Le cahier des charges traduit la décision de lancement de l'évaluation dans des termes suffisamment précis pour permettre la mise en concurrence de plusieurs équipes d'évaluation et pour garantir la réalisation du travail dans de bonnes conditions. La rédaction du cahier des charges mérite donc beaucoup de soin.

Le cahier des charges récapitule l'origine, le champ et la finalité de l'évaluation ainsi que la répartition des rôles. Il décrit le contenu du programme à évaluer et précise l'information qui est déjà disponible et qui sera mise à disposition de l'équipe d'évaluation. Il énonce, de façon précise, un nombre limité de questions auxquelles l'évaluation devra répondre. Il trace les grandes lignes de la méthode d'évaluation et fixe les étapes du calendrier de travail. Il indique éventuellement l'enveloppe budgétaire disponible

L'équipe d'évaluation est choisie après mise en concurrence. Parmi les critères à considérer, l'indépendance et la compétence sont deux éléments essentiels.

■ Soigner la rédaction du cahier des charges

Le cahier des charges (ou termes de référence) est le document servant de base à une relation contractuelle entre le commanditaire d'une évaluation et l'équipe chargée de l'étude. Construire de façon claire le cahier des charges est une étape vitale lorsque l'évaluation doit être réalisée par des consultants extérieurs. Ce travail revêt néanmoins une importance toute aussi grande si une partie de l'évaluation est faite en interne bien que l'on croie souvent pouvoir en faire l'économie dans ce cas. Le cahier des charges peut concerner soit l'opération d'évaluation dans son ensemble (lorsqu'elle est confiée à une équipe unique), soit un sous-ensemble de travaux d'étude programmés dans le projet d'évaluation (analyse approfondie d'une question évaluative).

Le cahier des charges est un texte de 5 à 20 pages, généralement complété par des annexes administratives. Son contenu est décrit dans l'Encadré 14. Seuls les points qui n'ont pas déjà été traités sont expliqués par la suite.

Exemple : une évaluation comprenant plusieurs cahiers des charges

Dans le cadre de l'évaluation de leur programme de développement, les autorités d'une région ont choisi de découper les travaux en lots séparés selon les capacités techniques requises pour les réaliser : (1) travaux d'enquête par questionnaire auprès d'un échantillon de destinataires, (2) travaux d'entretiens avec les gestionnaires et les opérateurs, (3) expertise économique. D'autres évaluations sont décomposées par territoire (ex : développement rural / développement urbain) ou par secteur d'intervention (ex : technologie / formation / tourisme). On notera qu'un tel découpage fait porter au commanditaire la responsabilité complète des choix méthodologiques et de la synthèse des résultats.

Encadré 14 - Le plan type d'un cahier des charges

- 1 - Cadre réglementaire
- 2 - Champ de l'évaluation
- 3 - Qui va utiliser l'évaluation et pour quoi ?
- 4 - Questions évaluatives et critères
- 5 - Connaissances et données disponibles
- 6 - Méthode (lignes directrices)
- 7 - Dispositif d'évaluation
- 8 - Calendrier
- 9 - Budget indicatif
- 10 - Qualifications requises
- 11 - Structure de la proposition

● **Cadre réglementaire de l'évaluation**

Pour le bénéfice de l'équipe qui sera chargée de réaliser l'évaluation, il est nécessaire de rappeler le contexte dans lequel celle-ci s'insère. Les bases juridiques et contractuelles sont explicitées dans les réglementations de l'Union européenne et dans chaque document de programmation mais il est souhaitable d'y faire référence.

Le cahier des charges doit préciser qui a été à l'initiative du projet d'évaluation et le cas échéant, qui a été associé à l'élaboration de la commande de cette évaluation et les motifs à l'origine de la décision d'évaluer s'ils ne sont pas purement réglementaires.

Par ailleurs, les finalités de l'évaluation, c'est-à-dire la raison pour laquelle une évaluation est demandée, doivent être explicitées : veut-on infléchir

une tendance ? pourquoi ? veut-on modifier les procédures de pilotage ou d'arbitrage ? veut-on réallouer des crédits ?

● Connaissances disponibles

Le cahier des charges contient un bilan de l'état actuel des connaissances sur le programme et ses effets. Parmi les éléments de cet état des lieux, on trouve des extraits ou des références des documents de programmation, la liste des études et des évaluations précédentes accompagnées d'extraits significatifs, la description du système de suivi, des indicateurs effectivement quantifiés, des divers comptes-rendus et bases de données disponibles auprès des services gestionnaires.

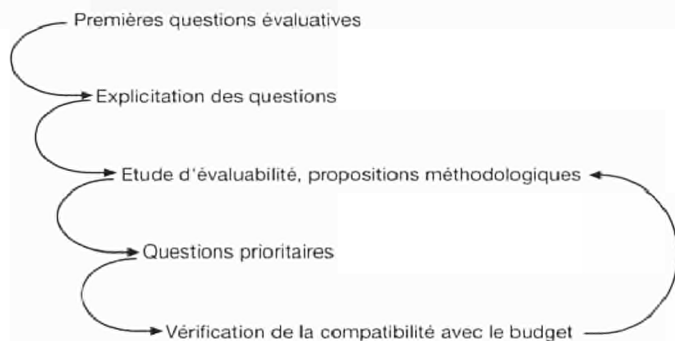
Cet état des lieux est indispensable pour permettre aux équipes candidates pour l'évaluation d'ajuster leurs propositions méthodologiques en fonction de l'ampleur et de la nature de la collecte de données et des travaux d'interprétation qui devront être organisés au cours de l'exercice d'évaluation.

● Grandes lignes de la méthode

Chaque évaluation fait appel à une méthode ad hoc construite avec l'accord du commanditaire, sur la base de la proposition de l'équipe d'évaluation sélectionnée. Au stade du cahier des charges, on choisit généralement de ne pas décrire complètement la méthode, mais de tracer les grandes lignes d'une méthode selon le schéma présenté à l'Encadré 15. L'objectif n'est pas d'enfermer les équipes candidates dans des contraintes méthodologiques strictes, mais d'énoncer quelques impératifs que le commanditaire considère comme essentiels pour obtenir une réponse à ses questions. Faut-il faire un inventaire des évaluations et rapports de recherche sur des sujets similaires ? Y aura-t-il enquête auprès des destinataires ou non ? Y aura-t-il une analyse comparative ou non ? L'équipe d'évaluation devra-t-elle rechercher des références de jugement ou ces références seront-elles fournies par le groupe de pilotage ?

On choisit généralement de garder une flexibilité suffisante pour permettre aux équipes candidates de se différencier selon la pertinence et la clarté de leurs propositions méthodologiques.

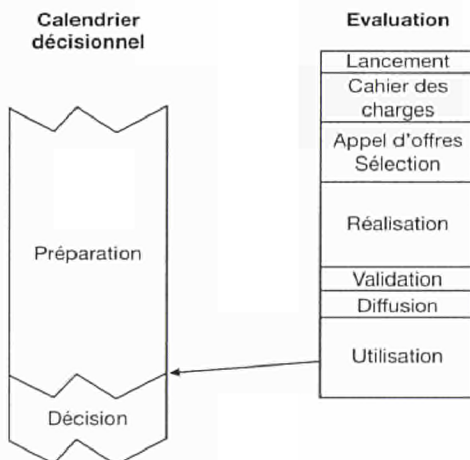
Encadré 15 - Prévoir une méthode d'évaluation au stade du cahier des charges



● Calendrier

Le calendrier de l'évaluation est établi en tenant compte de diverses contraintes dont les principales sont celles du calendrier décisionnel. Comme le montre l'encadré ci-dessous, il faut également intégrer les principaux délais incompressibles qui sont engendrés par les procédures d'appel d'offre et par les phases de collectes de données primaires.

Encadré 16 – Les contraintes de calendrier d'une évaluation



● Budget indicatif

Les ressources financières disponibles pour une étude déterminent en partie les limites de l'évaluation, et en particulier le nombre des questions traitées et l'approfondissement des traitements. Lorsque cela est possible, le budget de l'évaluation devrait être établi par le commanditaire en considérant l'intérêt et la difficulté des questions posées. Pour établir un budget indicatif, on peut adopter une démarche « descendante », une démarche « ascendante » ou une combinaison des deux.

Une première démarche « descendante » consiste à fixer le budget de l'évaluation par référence au budget du programme à évaluer. Quelques références peuvent être utilisées dans cette démarche. Une étude sur des évaluations réalisées dans le cadre du CCA irlandais pour la période 1994-1999 a montré que le coût annualisé de l'évaluation s'élevait entre 0,6 et 1,4% des dépenses annuelles du Programme Opérationnel (Etude du processus d'évaluation à mi-parcours dans le CCA irlandais pour la période 1994-1999, publié en 1998). Ces chiffres sont bien supérieurs à la dépense actuellement consacrée à l'évaluation des Fonds structurels européens, qui est globalement de l'ordre de 0,02% des fonds programmés. Dans certains cas, le coût de l'évaluation s'est élevée à 10%, ou même à 50% du coût de l'action évaluée. De tels pourcentages sont exceptionnels et se justifient pour des interventions très innovantes ou pour des opérations expérimentales. Dans ces circonstances, il n'est pas pertinent de se référer au coût de l'intervention évaluée. Une meilleure référence est le coût du programme qui va être décidé si l'évaluation conclut au succès de l'expérimentation.

Les facteurs qui peuvent justifier un budget d'évaluation faible ou élevé sont notamment les suivants :

- il s'agit d'une grande politique onéreuse, simple et routinière (% réduit) ;
- le programme est évalué fréquemment (% élevé pour l'ensemble des évaluations, mais réduit pour une évaluation individuelle) ;
- il s'agit d'une politique réglementaire dont les effets sont importants et dont le coût est minime (% important) ;
- il s'agit d'un programme innovant ou expérimental (% important).

La démarche « ascendante », quant à elle, consiste à chiffrer le coût probable des études à réaliser pour répondre à chaque question évaluative. Le coût d'une évaluation comprend deux grandes catégories de dépenses : les coûts généraux (structuration de l'évaluation et établissement du

rapport) et les coûts spécifiques de collecte et d'analyse des données. Les coûts généraux comprennent les dépenses relatives à l'analyse des documents du programme et des données du suivi, au bilan de la documentation, aux entretiens et réunions avec les gestionnaires et les principales parties prenantes, à l'établissement, la présentation et les discussions des rapports intermédiaires et finals. Les coûts spécifiques concernent les études de cas, les entretiens de groupe, les entretiens individuels, les questionnaires, les analyses statistiques, etc. On peut estimer le coût et la durée de mise en œuvre de quelques techniques d'évaluation de la façon suivante :

Encadré 17 – Indications de coût de quelques techniques d'évaluation

	Coût (personne / mois)	Durée (mois)
Enquête par questionnaires	Moyen	XXX
Enquête par entretiens individuels	Moyen	XX
Enquête par entretiens de groupe	Faible	XX
Enquête par études de cas	Moyen	XXX
Analyse statistique	Faible	XX
Panel d'experts	Moyen	XXXX
Utilisation d'un modèle économétrique	Faible	XX
Développement d'un modèle économétrique	Fort	XXXXXX
X = environ un mois		

En réalité, le coût de l'évaluation d'un programme varie considérablement d'un cas à l'autre. Le budget indicatif est souvent fixé en combinant les démarches ascendante et descendante et en prenant en compte des facteurs significatifs tels que :

- Le nombre des questions à approfondir. En général, le coût de collecte et d'analyse des données est proportionnel au nombre de questions approfondies, mais il est quelquefois possible de regrouper certaines de ces études analyses.
- Les attentes en terme de crédibilité. Quoique cela ne soit pas une règle absolue, les méthodes d'évaluation les plus solides ont tendance à être les plus chères.

- La qualité du système de suivi et l'existence d'un enregistrement des réalisations et des résultats.
- Il est également important de ne pas sous-estimer le coût, souvent caché, du temps consacré aux réunions du groupe de pilotage et à la production de données internes à l'administration.

■ Sélectionner l'équipe d'évaluation

L'équipe d'évaluation est sélectionnée en tenant compte de son indépendance, de son profil et des qualités méthodologiques de sa proposition.

● Indépendance de l'équipe d'évaluation

L'indépendance de l'évaluation ou de l'équipe d'évaluation est une exigence dans un contexte partenarial. En particulier, on doit éviter que l'équipe d'évaluation soit dans une relation de subordination quelconque avec les gestionnaires du programme évalué. Ceci est particulièrement important s'il s'agit d'une évaluation à mi-parcours ou si l'évaluation a vocation à porter un jugement sur la réussite du programme. On verra cependant au dernier chapitre de cet ouvrage que le souci d'indépendance ne doit pas empêcher l'équipe d'évaluation d'établir des relations étroites avec les gestionnaires. Le principe d'indépendance doit également être appliqué avec discernement lorsqu'il est plus efficace, en termes de coûts, de confier une partie du travail aux administrations chargées du suivi du programme. C'est le cas, par exemple, s'il s'agit de retrouver certaines données de gestion et de réaliser un premier traitement. Dans certaines circonstances, des mélanges habiles de contributions internes et externes sont probablement plus productifs que de confier l'ensemble de l'exercice à des professionnels extérieurs. Cependant l'équipe d'évaluation devra être en mesure de contrôler l'impartialité et la qualité de toute l'évaluation.

L'indépendance de l'évaluation n'est pas totalement garantie par l'appel à une équipe externe. Les membres d'une équipe d'évaluation peuvent être soumis à des pressions multiples à l'occasion de l'établissement du plan de travail, de l'accès à l'information et de la validation du rapport. L'équipe d'évaluation n'a pas toujours suffisamment de poids pour résister aux pressions, particulièrement si elle est peu familiarisée avec l'évaluation, si elle ne peut pas s'appuyer sur des données solides, ou si elle n'a pas le temps de faire des analyses complètes. Dans tous les cas, le groupe de

pilotage peut jouer un rôle complémentaire très utile pour améliorer l'indépendance de l'évaluation, notamment s'il est élargi à un ou plusieurs experts extérieurs.

Certains programmes des Fonds structurels font l'objet d'une succession d'évaluations confiées à la même équipe, par exemple plusieurs analyses approfondies et une évaluation à mi-parcours. Le plus souvent, les travaux sont confiés à une société de consultants ou à un centre de recherche, mais il arrive également que des unités d'évaluation soient constituées en marge de l'administration nationale pour jouer ce rôle (cas de l'Irlande par exemple). Dans ces situations, l'équipe d'évaluation travaille pendant plusieurs années au contact des gestionnaires du programme. Que l'équipe soit externe ou interne, elle parvient d'autant mieux à préserver son indépendance qu'il existe un groupe de pilotage bien structuré et comprenant des experts extérieurs. Une solution élégante consiste à faire appel à des évaluateurs non ressortissants du pays considéré, même si les différences linguistiques limitent souvent ce genre de pratique.

● Profils des équipes d'évaluation

Dans le domaine des Fonds structurels européens, les organismes intervenant sur le marché de l'évaluation sont de plus en plus nombreux. Les prestataires choisis appartiennent à des cabinets de consultants locaux, nationaux ou internationaux. Le secteur marchand occupe la part la plus importante des interventions. La présence des centres de recherche universitaire est également significative. Chaque équipe d'évaluation a un profil propre selon son métier de base. Une différence importante est liée à ce qui distingue l'évaluation et la recherche.

La recherche, surtout lorsqu'elle est fondamentale, a pour vocation principale de faire « avancer » les connaissances. Elle a une finalité cognitive. En principe, sa première préoccupation n'est pas de faire changer des comportements, ou tout au moins sa démarche ne vise pas une utilité immédiate. Bien qu'ayant recours aux mêmes règles méthodologiques que celles utilisées par la recherche (en sociologie, économie, etc.) l'évaluation a au contraire pour vocation première d'être au service direct des acteurs et de porter un jugement. Elle a une finalité normative et instrumentale. Sa première vocation est d'être utile et utilisée. Cette distinction fondamentale entre les approches « recherche » et « évaluation » explique pourquoi, même avec un cahier des charges précis, s'adresser à un cabinet de

consultants ou à un laboratoire universitaire aura des répercussions en termes de démarche et de résultats de l'évaluation. Les universitaires présentent l'avantage d'être perçus comme indépendants et très crédibles de par leurs exigences académiques. Alors que les cabinets privés quant à eux sont souvent plus disponibles en termes de calendrier et ont un plus grand souci de répondre aux attentes du commanditaire.

Au moment de faire appel à des équipes candidates, le groupe de pilotage doit préciser les compétences qui sont attendues. En même temps qu'elle fixe les grandes lignes de la méthode, une étude d'évaluabilité peut définir ces compétences et précise si l'évaluation a besoin d'experts spécialisés dans les domaines d'intervention du programme (sciences régionales, macroéconomie, marché de l'emploi, développement rural, etc.) ou si l'on recherche plutôt une compétence généraliste en évaluation, ou encore une compétence méthodologique spécifique (enquête, audit, systèmes d'information, analyse de causalité, etc.).

D'une manière générale, chaque consultant ou universitaire possède des points forts dans certains domaines techniques de l'évaluation et des points faibles dans d'autres. Le regroupement au sein d'une équipe mixte constitue souvent une solution, sous réserve que la coordination soit en mesure de jouer son rôle. Si les équipes candidates sont suffisamment polyvalentes, il devient possible de les mettre réellement en concurrence et ne pas faire appel aux mêmes experts de façon répétée. Cette situation crée de meilleures possibilités d'indépendance.

● **Qualités méthodologiques de la proposition**

La sélection de l'équipe n'est jamais indépendante du critère de prix. Dans ce cas, il convient que la méthode d'évaluation soit déterminée avec beaucoup de détails et que les équipes candidates soient présélectionnées en fonction de leurs compétences techniques. Cette formule oblige le commanditaire et le groupe de pilotage à faire un travail méthodologique considérable en amont.

Lorsque la réglementation administrative le permet, il est préférable de procéder à l'inverse, c'est à dire de fixer une enveloppe de budget ou de volume de travail, de décrire seulement les grandes lignes de la méthode dans le cahier des charges, et de sélectionner l'équipe qui propose la méthode la plus intéressante. La sélection de l'équipe suppose alors que l'on sache juger la qualité méthodologique d'une offre. Cette qualité n'est

pas nécessairement liée à l'utilisation d'une terminologie obscure et de modèles compliqués. Il convient surtout de s'assurer que la méthode proposée permettra réellement de répondre aux questions posées dans le cahier des charges. Le groupe de pilotage peut conduire ce travail en vérifiant pour chacune des offres, question par question, les points présentés dans l'Encadré 18.

Encadré 18 - Examen de la qualité méthodologique d'une proposition

	<i>Pour chaque question d'évaluation</i>		
	Question 1	Question 2	...
<i>La proposition de l'équipe candidate :</i>			
<i>Prévoit-elle de rassembler et/ou de collecter des informations suffisamment pertinentes ?</i>	++	+	
<i>Se fonde-t-elle sur des techniques rigoureuses d'analyse ?</i>	-	+	
<i>Se donne-t-elle les moyens d'explicitier les critères d'évaluation de façon impartiale ?</i>	+	±	
<i>Produira-t-elle vraisemblablement des conclusions crédibles ?</i>	+	+	
<i>L'importance respective des questions a-t-elle été bien appréhendée ?</i>	++		

III

Réaliser l'évaluation d'ensemble



Analyser la logique et le contenu du programme

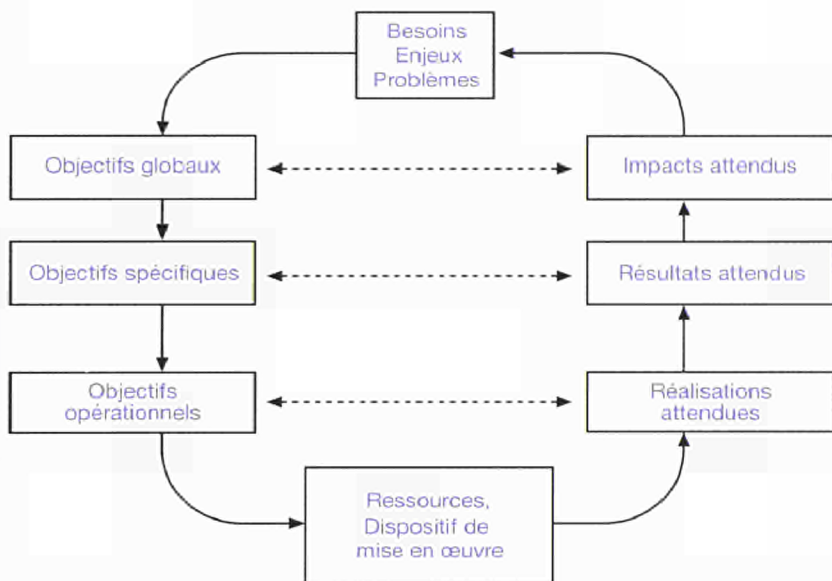
Dans un premier temps, l'équipe d'évaluation reconstitue la logique d'un programme, c'est à dire les besoins à satisfaire, la stratégie choisie, les objectifs fixés, les effets attendus, les ressources allouées, ainsi que les dispositions prises pour assurer la mise en œuvre et le suivi. L'analyse de la logique du programme construit un cadre indispensable pour l'étude des résultats et des impacts. C'est donc une étape clé de toute évaluation.

Lors d'une évaluation *ex ante*, cet examen constitue l'essentiel du travail car il débouche sur un jugement du bien fondé et de la cohérence du programme et de sa stratégie. Lors d'une évaluation à mi-parcours, cet examen débouche sur des recommandations d'amélioration du système de suivi et des méthodes de gestion comme sur des propositions d'ajustement des programmes.

Dans le cadre d'une évaluation *ex post*, cet examen permet de clarifier, si nécessaire, des objectifs qui n'avaient pas été clairement exprimés lors des étapes précédentes. Cette clarification est indispensable pour structurer au mieux la collecte d'informations et l'analyse des impacts.

De façon générale, cette étape va permettre de vérifier la cohérence des objectifs entre eux et avec ceux d'autres interventions. Le contenu du programme sera examiné au regard des ressources budgétaires allouées, des modes d'intervention choisis (sélections des projets, responsabilités des intervenants), des impacts affichés et du système de suivi mis en place au sein duquel le système d'indicateurs a une place importante.

La logique d'un programme se définit comme l'articulation entre tous ses éléments, à savoir : les besoins à satisfaire, la stratégie choisie, les objectifs fixés (globaux, spécifiques, opérationnels), les ressources allouées, les dispositions prises pour la mise en œuvre, et les effets attendus (réalisations, résultats et impacts). On a coutume de présenter l'articulation de tous ces éléments sous la forme d'une succession logique d'étapes dénommée « cycle de programmation », telle que figurée dans l'encadré 19 ci-après.

Encadré 19 – Le cycle de programmation

◀ L'évaluation ex ante examine la logique du programme proposé et porte un jugement sur le fait qu'elle est claire, complète et cohérente. Il s'agit de sa vocation principale. Les pages qui suivent constituent donc une étape clé de la réalisation d'une évaluation ex ante. Dans une seconde étape, l'évaluation ex ante réalise une estimation prévisionnelle des impacts.

▶▶ L'évaluation à mi-parcours et l'évaluation ex post peuvent aussi consacrer leur première étape à la logique du programme, comme le montre le schéma de l'Encadré 19. Cela est évidemment nécessaire si la logique de départ présente des points obscurs ou mal justifiés, c'est à dire si l'évaluation ex ante n'a pas été conduite correctement ou si ses conclusions n'ont pas été entendues. Il peut également arriver que les objectifs aient été modifiés ou aménagés en cours de route, ce qui impose de présenter la nouvelle logique du programme au début du rapport d'évaluation et d'examiner si les changements d'objectifs ont préservé la cohérence d'ensemble. Plus fréquemment, il arrive que les besoins à satisfaire ou les problèmes à résoudre aient changé pendant la mise en

œuvre du programme. Si nécessaire, l'évaluation à mi-parcours ou l'évaluation ex post réexaminent alors la logique du programme à la lumière de ce nouveau contexte.

Lorsque l'équipe d'évaluation examine la logique du programme, elle commence par analyser les textes officiels (proposition de programme dans le cas d'une évaluation ex ante, programme approuvé dans les autres cas). Il s'agit alors d'explicitier la « logique officielle » et d'évaluer sa cohérence. Si les autorités gestionnaires du programme le souhaitent, l'évaluation peut se poursuivre par une série d'entretiens et/ou de réunions au cours desquelles les participants expliquent comment ils comprennent la stratégie, les objectifs et les effets attendus. Il s'agit alors de produire, de façon participative, « une logique reconstruite ». La première option est plus rapide et elle a l'avantage de respecter les textes officiels, ce qui permet de rendre des comptes sur une base légitime. L'inconvénient est que l'évaluation peut être amenée à révéler des incohérences ou des imprécisions qui sont déjà largement connues des décideurs. La seconde option nécessite plus de travail mais elle permet de produire des conclusions plus utiles pour l'amélioration du programme.

■ Apprécier le bien-fondé du programme

La raison d'être d'un programme est de satisfaire des besoins, de résoudre des problèmes de prévenir des menaces ou de saisir des opportunités. Les besoins ou les problèmes considérés sont de nature sociale, économique ou environnementale. Ils concernent un territoire ainsi qu'un ou plusieurs publics sur ce territoire. S'il s'agit de justifier un financement public, les besoins à prendre en compte sont ceux qui ne peuvent pas être satisfaits par le marché et les problèmes sont ceux qui ne peuvent pas être résolus par l'initiative privée.

● Retour sur le diagnostic socio-économique

◀ L'équipe d'évaluation a comme première tâche incontournable de rappeler les besoins sociaux et les problèmes de développement qui justifient l'existence du programme. Elle doit examiner la logique de cette justification de façon complète et systématique, particulièrement dans le cas d'une évaluation ex ante.

L'évaluation précise quels sont les handicaps socio-économiques et les problèmes de développement qui sont rencontrés par la région ou par

l'Etat membre considéré, et qui justifient qu'une aide soit allouée au titre des Fonds structurels. L'équipe d'évaluation se prononce sur le fait que les besoins invoqués correspondent effectivement aux objectifs généraux de la politique européenne de cohésion.

► Dans le cas d'une évaluation à mi-parcours, l'équipe d'évaluation devra également examiner l'évolution que connaissent ces besoins socio-économiques ou ces problèmes de développement. Elle regardera s'ils vont vers une aggravation ou une résorption et se demandera pour quelles raisons. L'objet de ce questionnement est en particulier de savoir si l'initiative privée pourrait satisfaire ces besoins ou résoudre ces problèmes à la place de et aussi bien que le programme public considéré.

Il faut garder à l'esprit comme l'illustre l'exemple ci-dessous que, selon les périodes considérées, certains problèmes peuvent être requalifiés comme relevant plus ou moins d'une intervention publique.

Exemple : un nouvel examen de la raison d'être des interventions

Dans le cadre de l'évaluation à mi-parcours de l'ensemble des programmes cofinancés par l'Union européenne en Irlande, l'équipe d'évaluation a classé les interventions en quatre grandes catégories selon la façon dont elles peuvent être justifiées : celles qui produisent des biens publics qui ne peuvent pas être fournis par l'initiative privée, celles qui corrigent des défaillances du marché, les subventions spécifiques et les mesures de redistribution. Au sein de chaque catégorie, l'équipe d'évaluation a noté le bien-fondé de chaque intervention à l'aide d'une même grille d'appréciation. Sur la base de cette évaluation, il est apparu que certaines mesures pouvaient être considérées comme ayant perdu leur raison d'être : parmi celles-ci, on retrouve certaines incitations à destination du secteur touristique privé qui devenaient inutiles en raison de la rapidité de la croissance. Au contraire, d'autres mesures avaient une pertinence renforcée, par exemple l'amélioration des routes inter-urbaines et la lutte contre l'abandon précoce de l'école.

● Pertinence de la stratégie et des objectifs

◀ Etablir la stratégie d'un programme revient à choisir des priorités d'action en fonction de l'urgence des besoins à satisfaire, de la gravité des problèmes à résoudre, et des chances de succès des actions envisageables. La formulation d'une stratégie aboutit à sélectionner les objectifs, à fixer leur niveau d'ambition et à les hiérarchiser. Tous les territoires ou les publics ne relèvent pas d'une même stratégie de

développement. L'évaluation *ex ante* examine si la stratégie est adaptée au contexte et à son évolution probable. Il convient donc de vérifier que le programme a bien été construit sur la base d'un classement des objectifs selon leur degré d'importance.

L'évolution du contexte est caractérisée par des enjeux auquel la stratégie doit répondre. Celle ci pourra alors prendre une forme active ou réactive selon la nature de l'enjeu. Il pourra s'agir par exemple d'une menace qu'il conviendra de prévenir, d'un possible blocage économique qui devra être écarté, d'une faiblesse dont il faudra tenir compte ou alors d'une opportunité qu'il faudra saisir, d'un potentiel de développement peu exploité qu'il faudra stimuler, d'une force sur laquelle on pourra s'appuyer, etc. La stratégie est composée d'objectifs défensifs et offensifs. Dans le premier cas, le programme visera plutôt à lutter contre une faiblesse ou à prévenir une menace. Dans le deuxième cas, il visera à accompagner une force ou à profiter d'une opportunité.

L'évaluation considère la stratégie du programme en cherchant à savoir si elle est principalement offensive ou défensive, ou éventuellement mixte. Une stratégie purement défensive comporte le risque de devoir durablement affronter de nouveaux problèmes en instituant l'idée que le territoire ou le public concerné ne peut se développer qu'avec un soutien extérieur. Une stratégie purement offensive peut conduire à négliger des menaces importantes et potentiellement catastrophiques pour le développement.

Une technique utilisable pour l'évaluation de la stratégie du programme peut être l'analyse SWOT. L'analyse SWOT (*Strengths-Weaknesses-Opportunities-Threats*), soit Forces-Faiblesses-Opportunités-Menaces, avec un groupe composé du maximum de parties prenantes concernées procure en effet une structure formelle qui permet de clarifier et de préciser en détails les forces et les faiblesses d'une situation donnée (régionale ou sectorielle) et les opportunités et les menaces existantes ou potentielles. L'outil est décrit plus en détail dans le Volume 3 de la Collection MEANS.

L'équipe d'évaluation examine également quel est le degré d'ambition de la stratégie par rapport aux besoins et problèmes identifiés. Il s'agit de montrer clairement l'esprit dans lequel le programme a été conçu : pour satisfaire complètement les besoins ou régler définitivement les problèmes au cours de son cycle de vie, ou comme la première étape d'un effort à prolonger sur une période plus longue, ou encore comme le test d'une solution à généraliser ultérieurement. Le degré d'ambition du programme

s'apprécie également par rapport à d'autres programmes comparables. Il est important que l'évaluation examine l'ambition des objectifs pris un par un. Ce travail est inséparable de l'évaluation de l'atteinte des objectifs car il est facile d'atteindre des objectifs peu ambitieux. L'évaluation de l'efficacité du programme doit être nuancée en conséquence.

Exemple : une analyse SWOT

L'analyse SWOT a été utilisée lors de l'élaboration du plan de développement régional pour l'Objectif 6 en Suède. Les résultats produits grâce à cette technique ont permis de mettre en évidence les grandes tendances environnementales, structurelles et démographiques caractéristiques du contexte de la région retenue. Les facteurs majeurs identifiés lors de cette analyse sont synthétisés ci-dessous :

Forces : Large espace naturels préservés, ressources naturelles importantes.

Faiblesses : Région isolée, faible densité de population, vulnérabilité aux fluctuations économiques, à celles des budgets publics, faiblesses structurelles des entreprises.

Opportunités : Tourisme, qualité des ressources forestières, production piscicole en eau douce, prospection minière.

Menaces : Déclin démographique, insuffisance du nombre d'emploi.

Les recommandations issues de cette analyse ont permis d'orienter les choix faits dans le cadre du Plan de Développement Régional. Cet outil a, depuis lors, été utilisé dans de nombreuses régions.

■ Vérifier la clarté et la cohérence des objectifs

● Clarté des objectifs

Lorsqu'un objectif a été clairement identifié, il est plus facile de constater en fin de période s'il a été atteint ou non. Un objectif parfaitement clair est donc un objectif vérifiable. Dans le premier des exemples qui suivent, l'objectif est relativement facile à vérifier. Il s'agit d'augmenter la dynamique de création d'entreprises. L'objectif est associé à un indicateur mesurable dont la signification est claire pour tous (taux de création d'entreprises pour 10 000 habitants). Il a été fixé sous la forme d'une valeur quantitative (+ 20%). Les exemples suivants montrent des objectifs quantifiés mais plus difficilement vérifiables, ainsi que des objectifs non quantifiés, dont certains sont plus faciles à vérifier que d'autres. Il revient à l'équipe d'évaluation de porter un jugement systématique sur la clarté des objectifs.

Exemples : des objectifs vérifiables et non vérifiables

Des objectifs facilement vérifiables

A la fin du programme, le ratio de création d'entreprises pour 10.000 habitants aura augmenté de 20%.

à la fin du programme, les principales villes de la région seront classées à un rang plus élevé par les sondages qui posent la question « où aimeriez vous habiter ? »

Des objectifs difficilement vérifiables

Le programme contribuera à 1% d'augmentation du PIB/habitant

Le programme renforcera la compétitivité des entreprises régionales

Lorsqu'un objectif est difficilement ou même impossible à vérifier, l'équipe d'évaluation se trouve devant un travail complémentaire à faire avant de poursuivre vers les étapes suivantes de l'évaluation.

◀▶ Dans le cas d'une évaluation ex ante ou à mi-parcours, l'équipe d'évaluation collecte toutes les opinions pertinentes afin de faire des recommandations dans le sens d'une clarification.

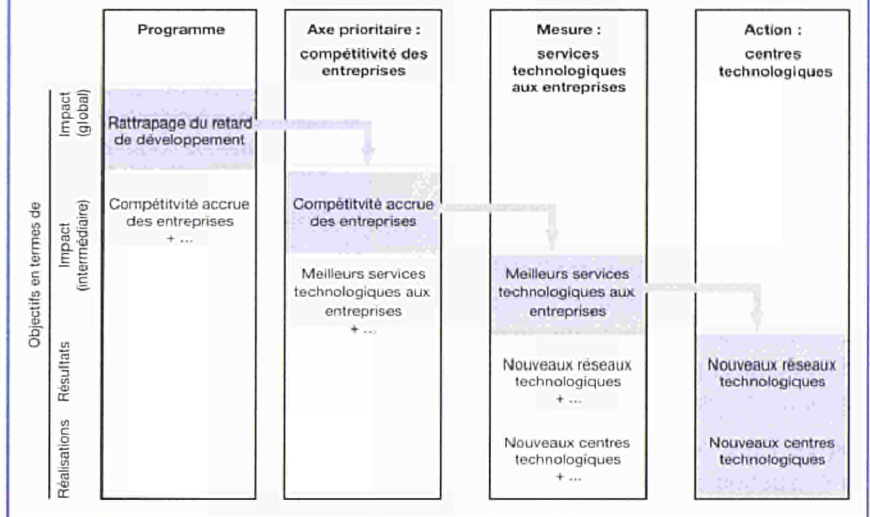
▶▶ Dans le cas d'une évaluation à mi-parcours ou ex post, elle recherche si les gestionnaires du programme ou les opérateurs ont redéfini voire clarifié au cours de la mise en œuvre les objectifs initiaux. Les entretiens devront permettre de savoir s'il y a consensus sur l'interprétation de la logique du programme.

Chercher à rendre un objectif mesurable à l'aide d'un indicateur aide souvent à le rendre plus clair.

● Cohérence des objectifs au sein d'un programme

L'évaluation vérifie que les objectifs du programme sont cohérents c'est à dire que les objectifs des mesures découlent logiquement des objectifs spécifiques définis pour les axes ou sous-programmes. A ce niveau, on parle de cohérence interne du programme ; celle ci étant définie comme une organisation des objectifs qui est telle que les objectifs de niveau inférieur constituent une contribution logique aux objectifs de niveau supérieur. C'est ce que l'on entend par la notion de hiérarchie des objectifs qui est représentée par le schéma ci-après.

Encadré 20 - Hiérarchie des objectifs



Adapté de : COMMISSION EUROPEENNE (1995) - Orientations communes pour le suivi et les évaluations intermédiaires - Fonds structurels

Pour vérifier la cohérence interne du programme, on peut avoir recours à la technique de l'arbre d'objectifs. En effet, cette technique permet de déterminer la contribution de chacun des objectifs de niveau inférieur à l'objectif global du programme. Il s'agit d'un schéma où l'objectif le plus global du programme est placé tout à gauche et les objectifs les plus détaillés tout à droite. Chaque objectif global peut ainsi se décomposer en plusieurs sous-objectifs qui, à leur tour, se décomposent en objectifs d'un niveau plus spécifique. Cet outil permet ainsi d'identifier et hiérarchiser les objectifs.

En pratique, ni les objectifs de chaque mesure (objectifs spécifiques), ni les liens entre les différents niveaux d'un programme complexe, ne sont généralement articulés aussi clairement et de façon aussi explicite. Le besoin de clarifier les différents niveaux d'objectifs d'un programme est alors d'autant plus nécessaire que l'on n'est pas certain des impacts de chaque type d'interventions.

L'arbre d'objectifs permet de spécifier les objectifs et d'établir une hiérarchie claire entre eux en identifiant clairement le degré de cohérence interne du programme, à la fois verticalement, entre les différents niveaux d'objectifs, et horizontalement, entre les objectifs de même niveau des

différentes mesures (Voir le Volume 4 de la Collection MEANS ainsi que le Volume 5 pour de nombreux exemples thématiques).

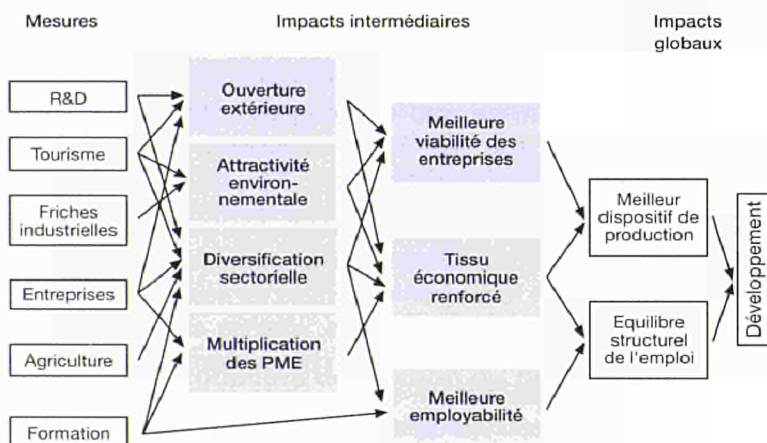
Exemple : objectifs cohérents et incohérents

Deux mesures visant à « offrir de meilleurs services technologiques aux PME » et à « offrir des possibilités de financement en capital risque aux PME » ont des objectifs cohérents avec l'objectif de niveau supérieur « accroître la compétitivité des entreprises ».

Au contraire, ces objectifs seraient incohérents si l'objectif de compétitivité était énoncé sans qu'aucune mesure n'ait contribué clairement ou si les mesures de niveau inférieur étaient juxtaposées sans rien avoir en commun.

Tandis que l'arbre d'objectifs permet d'identifier et de hiérarchiser les objectifs, il ne permet pas d'identifier les liens de causalité qui existent entre ceux-ci. Pour ce faire, on a recours à un autre outil le « diagramme logique des impacts attendus ». Cet outil schématise les étapes de causalité entre les réalisations du programme, situées à gauche, et les impacts spécifiques jusqu'à l'impact global ou aux impacts globaux, situés à droite. L'arbre d'objectifs s'inscrit dans une logique de décomposition d'un objectif global, par contre le diagramme logique d'impacts s'inscrit dans une logique de reconstitution d'impacts attendus ou d'objectifs spécifiques (Voir Encadré 21 et Encadré 41).

Encadré 21 – Exemple de diagramme logique des impacts attendus



● Cohérence des objectifs avec les autres politiques publiques

Il est important de se rappeler que le programme ne représente qu'une partie des politiques publiques menées dans une région donnée et que les ressources qui lui sont allouées ne sont qu'une petite part des ressources publiques dépensées localement. Certains des besoins que le programme peut légitimement chercher à couvrir constituent également la cible d'autres politiques, qu'elles soient régionales, nationales ou européennes. Par conséquent, on doit s'assurer que les activités programmées ne sont pas en conflit ou ne viennent pas dupliquer inutilement ces autres politiques. Il s'agit donc d'apprécier la cohérence externe du programme, c'est à dire d'examiner notamment si les objectifs du programme ne sont pas en contradiction avec les autres politiques de l'Union européenne, par exemple celles relatives à la concurrence ou à l'organisation des marchés agricoles.

Exemple : la cohérence externe dans le cadre d'interventions pour le développement rural

Les conditions d'attribution de certaines aides de la Politique Agricole Commune dans le cadre de l'organisation des marchés agricoles, visant au maintien de l'espace rural, sont en cohérence avec les objectifs des programmes de développement économique des zones rurales financées par les Fonds structurels. Pour éviter toute duplication, il est important que les programmes 5b soient entièrement consacrés au maintien d'une communauté agricole viable dans les régions en difficulté : maintien de la population dans les zones rurales défavorisées, maintien et développement de l'activité agricole dans ces mêmes zones.

◀ La réglementation communautaire précise par ailleurs qu'une information sur la cohérence avec les politiques économiques, sociales et, le cas échéant, régionales des Etats membres, doit être inscrite dans chaque document de programmation.

▶ Le principal risque qui pèse sur la cohérence externe est l'apparition, en cours de route, d'autres interventions qui font double emploi ou qui créent des contradictions avec le programme évalué. Si nécessaire, l'évaluation à mi-parcours fait un bilan des politiques voisines et estime dans quelle mesure elles prolongent, complètent, chevauchent ou contredisent le programme en cours de mise en œuvre. Cette étude est menée à l'aide d'entretiens avec les gestionnaires des programmes, des opérateurs et de tout autre informateur compétent.

Une des techniques qui peut être utilisée pour clarifier les relations avec les différentes autres politiques est d'établir une matrice des impacts croisés (composantes du programme en ligne et autres politiques en colonnes). Une telle matrice va montrer les complémentarités, les duplications et les conflits qui risquent de surgir entre les impacts du programme considéré et les autres politiques existantes (Voir Volume 4 de la Collection MEANS). L'utilisation de cet outil est montrée dans l'Encadré 22. Dans cet exemple on voit qu'il existe de nombreux cas de synergie entre les activités programmées et les autres politiques mise en œuvre, par exemple, au niveau national. A titre d'exemple, « l'appui à la diversification » et « la promotion des investissements étrangers » se renforcent. Par contre, cette dernière politique qui s'appuie sur des ressources exogènes est en conflit avec un programme en faveur du développement endogène soutenu par la création d'un réseau d'appui pour les PME.

Encadré 22 - Mise en évidence de la cohérence externe à l'aide d'une matrice d'impacts croisés

Activités programmées	Autres politiques		
	Promotion régionale des investissements étrangers	Promotion nationale des exportations	Promotion européenne des PME
Appui à la diversification	+	+	?
Appui aux investissements productifs	?	+	+
Réseau d'appui	-	+	+
Appui aux services aux entreprises	=	+	+
Renforcement régional de la R&D	+	=	=
Promotion d'une nouvelle image de la région.	+	=	=
= : pas d'interaction + : complémentarité ? : risque de duplication - : risque de conflit			

■ Analyser les modalités de gestion du programme

Le processus de mise en œuvre suppose de prendre une série de dispositions organisationnelles et budgétaires à savoir : l'établissement du

budget, la sélection des projets, les circuits financiers, les modalités de gestion et de coordination, le système de suivi et de comptes-rendus.

A ce stade de l'évaluation, on examine les dispositions budgétaires, les consignes opérationnelles données, ou encore les guides diffusés pour faire en sorte que le programme soit mis en œuvre avec le maximum de chances d'atteindre ses objectifs. L'analyse proprement dite de l'adéquation des conditions de mise en œuvre sera faite à un stade ultérieur du processus d'évaluation notamment lorsque l'on recherchera les raisons de l'atteinte ou non d'un impact donné.

◀ Dans le cadre d'une évaluation ex ante, l'examen se limitera aux premières dispositions décrites dans la proposition de programme, c'est à dire principalement les enveloppes budgétaires, les instruments choisis et les principales règles du jeu établies.

▶ Dans le cadre des évaluations à mi-parcours ou ex post, il est possible d'aller beaucoup plus loin et d'examiner également le détail des dispositions administratives, la répartition des rôles entre gestionnaires et opérateurs, les documents d'application transmis aux opérateurs, les critères de sélection des projets, etc.

● Répartition des ressources budgétaires

L'évaluation s'attache principalement aux financements publics alloués par les autorités qui ont « co-décidé » le programme. Les investissements privés complémentaires ne sont pas à mettre sur le même plan, de même que les fonds publics en provenance d'autres financeurs dont la contribution est espérée mais pas formellement programmée. Le programme répartit ses ressources budgétaires sous forme d'enveloppes allouées à ses différentes composantes. Avant même d'examiner comment les ressources sont attribuées aux agences et aux autorités locales chargées de la mise en œuvre, l'évaluation peut examiner la cohérence entre l'allocation des ressources et les objectifs. Elle examinera également l'effet potentiel de levier ou d'entraînement que ces ressources publiques auront sur la mobilisation d'autres fonds, en particulier privés. Sur ce dernier point voir le Volume 5 de la Collection.

◀ L'examen porte de façon plus détaillée sur la correspondance entre la répartition des ressources budgétaires et les priorités stratégiques du programme. L'évaluation vérifie que l'importance relative des enveloppes

budgétaires attribuées aux différents objectifs reflète bien les priorités et les ambitions annoncées. A l'échelle du programme, il s'agit d'appliquer le même principe qui s'applique pour l'ensemble de la politique européenne de cohésion sous le nom de concentration. Ce principe exprime que la majeure partie des ressources doit être dirigée vers l'atteinte des objectifs considérés comme prioritaires.

La concentration s'évalue en répondant à des questions du type de celles qui suivent : dans quelle mesure le programme s'est-il concentré sur quelques objectifs stratégiques ? sur des zones restreintes devant faire face à des problèmes particulièrement difficiles ? ou sur un petit nombre de publics prioritaires ?

Pour réaliser cette étape de l'évaluation, on peut utiliser l'arbre des objectifs s'il en a été dessiné un à l'étape précédente. On inscrira simplement sur chaque branche de l'arbre le montant total des ressources attribuées à l'objectif correspondant.

● Modes d'intervention

Il y a souvent plusieurs façons d'atteindre un objectif. Chacune de ces façons correspond à différents instruments qu'il est possible d'employer, par exemple les subventions directes, le conseil, la formation, la fourniture de services gratuits, la construction d'infrastructures, etc.

Dans le domaine des ressources humaines, d'importants financements sont attribués à des programmes ou à des systèmes de formation et d'insertion professionnelle. Les destinataires sont touchés par des organismes de formation ou par l'intermédiaire de leurs entreprises. Les programmes d'insertion fournissent une aide aux organismes publics ou privés qui embauchent des chômeurs de longue durée. D'autres programmes favorisent par exemple la cessation d'activité pour les agriculteurs âgés au profit de jeunes agriculteurs qui souhaitent s'installer.

Dans le domaine de l'environnement productif, l'aide peut être directe (cas d'une subvention à l'investissement ou d'un prêt bonifié) ou indirecte (cas d'un pôle d'innovation). Elle peut prendre la forme d'un régime de subvention ouvert ou d'un programme plus sélectif. L'aide peut être réservée à des investissements physiques ou à des investissements immatériels (marketing, prospection internationale, réseaux, innovation, qualité, design, gestion, etc.). Elle passe souvent par une amélioration de

l'offre de services aux entreprises. Les programmes de recherche - développement représentent la forme la plus indirecte de soutien aux structures productives.

◀ Dans le domaine des infrastructures de base, l'aide est décidée dans le cadre de programmes ou de grands projets. Certains projets comportent un accès payant et d'autres non. Certains projets sont réalisés par des autorités publiques et d'autres sont délégués à des opérateurs privés.

Pour chaque objectif, l'évaluation peut examiner les instruments alternatifs qui sont utilisables pour l'atteindre et les raisons qui ont justifié le choix d'un instrument plutôt qu'un autre. Il est également possible de juger la pertinence du choix des instruments au regard de l'expérience connue par ailleurs.

● Critères de sélection des projets

Un programme est généralement constitué de mesures, lesquelles regroupent un certain nombre de projets qu'il faut retenir, c'est à dire choisir en fonction de critères à déterminer. Ces critères sont des critères de sélection qui doivent être établis avant le démarrage du programme.

▶ Par conséquent, dans le cadre de l'évaluation d'un programme, il sera nécessaire d'examiner si la gestion de celui-ci est clairement orientée vers l'atteinte des objectifs et pour ce faire, d'examiner si les projets ont été choisis judicieusement. Cela revient à examiner si les critères de choix des projets ont été correctement définis au regard des objectifs du programme (Voir Exemple, page 56).

Cette recherche de la cohérence entre les critères de sélection des projets et les objectifs du programme aura une grande importance dans le cas d'une évaluation à mi-parcours, à un moment où le programme peut encore être amélioré. En effet, dans ce cas, l'évaluation va contribuer aux choix des critères qui seront une des clés de la réussite du programme par le biais du type de projets à retenir ou à privilégier.

Deux types de questions devront être posés :

- Quels critères de sélection des projets sont employés ?
- Les critères de sélection sont-ils clairement en rapport avec les objectifs du programme ?

● Répartition des responsabilités opérationnelles

Les ressources qui sont à la base d'une intervention publique ne sont pas seulement financières. Elles sont également d'ordre humain et organisationnel ; ces deux composantes étant plus difficiles à appréhender avec précision et donc à estimer.

◀ ▶ Cependant, tout comme elle récapitule les ressources financières, l'évaluation va devoir examiner qui sont les acteurs chargés de la mise en œuvre du programme à chaque niveau (programme, mesures, projets) et comment est organisée l'articulation entre les différents niveaux et les différents acteurs.

L'évaluation va examiner si à chaque niveau, chaque objectif est clairement sous la responsabilité d'au moins un gestionnaire. Lorsqu'un objectif est sous la responsabilité de plusieurs gestionnaires, elle va examiner si un mécanisme de coordination a été mis en place.

Les programmes des Fonds structurels européens faisant l'objet d'une mise en œuvre partenariale, l'évaluation doit aussi se demander quelle en est la conséquence sur la cohérence au sein du programme (cohérence interne) et avec les autres politiques appliquées dans la région (cohérence externe). La démarche théorique d'évaluation applicable en la circonstance consiste à se demander et à estimer dans quelle mesure les effets du programme ont été meilleurs et plus importants grâce au partenariat. Cette approche permettra d'observer si les effets du programme entrent en synergie ou en contradiction avec ceux des politiques conduites dans la région par les autres partenaires.

Toujours concernant les différentes formes de ressources, l'évaluation devra rechercher s'il y a de gros problèmes d'efficacité résultant par exemple, de retards dans les circuits financiers, des processus de contrôle, des systèmes de décision, etc.

■ Examiner le système d'information du programme

Le « système d'information » du programme est l'ensemble des données rassemblées dans le cadre de la conception du programme puis collectées de façon permanente par les opérateurs au stade de la mise en œuvre (demandes d'aides, appréciation des projets au moment de la sélection, paiements des aides, comptes rendus d'avancement des projets, comptes rendus finals des projets, etc.). Une partie de ces données sont agrégées,

interprétées et transmises sous forme de tableaux ou de rapports de synthèse. Une autre partie de ces données sont conservées à des fins de contrôle a posteriori et peuvent aussi être utilisées par l'évaluation.

● **Examen des résultats attendus**

La décomposition de la logique du programme permet de distinguer plusieurs niveaux d'objectifs qui s'articulent entre eux. A chaque niveau d'objectif (opérationnels, spécifiques et globaux) sont associés respectivement des réalisations attendues, des résultats et des impacts (Cf. Encadré 19). Ces notions sont présentées plus en détail dans le Volume 2 de la Collection MEANS.

Les réalisations constituent le produit de l'activité des opérateurs. Ils représentent ce qui est obtenu en contrepartie de la dépense publique. Pour savoir si s'agit bien d'une réalisation, il suffit de poser la question suivante : « est-ce la condition pour le paiement du financement public ? ». Les réalisations sont sous l'entière responsabilité des opérateurs qui doivent en rendre compte de façon permanente par l'intermédiaire du système de suivi. On citera comme exemples de réalisations : les routes construites, les friches urbaines réhabilitées, les stations d'épuration construites, les formations qui ont été payées par le programme.

Les résultats représentent les avantages immédiats du programme (ou exceptionnellement les inconvénients immédiats) pour les destinataires directs. Un avantage est immédiat si le destinataire en bénéficie lorsqu'il est en interaction direct avec les réalisations du programme. Les résultats attendus d'un projet précis sont le plus souvent enregistrés dans le cadre du suivi car ils figurent généralement dans le dossier de demande d'aide déposé par l'opérateur ou le destinataire direct.

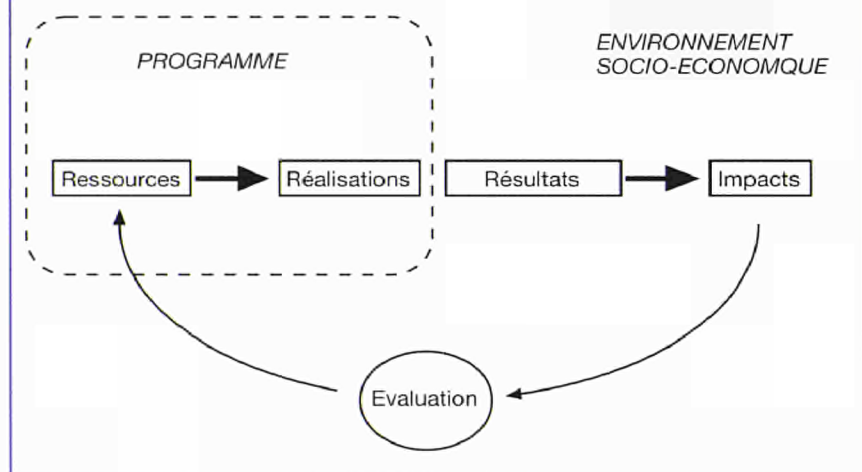
Les résultats correspondent à des changements qui interviennent pour les destinataires directs, par exemple : gain de temps des usagers d'une route, diminution du prix des communications téléphoniques, qualification acquises par des stagiaires, diversification de l'activité d'un agriculteur, nouvelle capacité de production installée par un industriel, satisfaction des entreprises qui ont reçu un conseil.

● **Identification des impacts attendus**

Comme représenté dans l'Encadré 23, les impacts représentent les conséquences qui résultent des interactions directes et immédiates du

programme avec ses destinataires. Une première catégorie d'impacts est donc constituée des conséquences qui apparaissent ou qui perdurent à moyen terme chez les destinataires directs du programme (impacts spécifiques). Par exemple, la circulation sur une route un an après son ouverture, le placement des stagiaires au bout de douze mois, les emplois durablement créés dans un établissement industriel construit avec l'aide du programme, la survie des entreprises créées avec l'aide du programme. Certains impacts peuvent être inattendus (retombées) mais ils ne font pas partie de la logique du programme.

Encadré 23 - Organisation publique : rétroactions par l'évaluation



Une seconde catégorie d'impacts est constituée de toutes les conséquences qui affectent, à court ou à moyen terme, des personnes ou organismes qui ne sont pas destinataires directs. Ces impacts peuvent être proches (par exemple : l'amélioration de la qualité de vie pour le voisinage d'une friche industrielle réhabilitée, l'amélioration de la qualité des plages au voisinage d'une nouvelle station d'épuration). Ils peuvent au contraire s'étendre à des personnes ou à des organisations très éloignées, comme dans le cas des impacts macroéconomiques.

Les mécanismes de propagation des impacts sont souvent distingués en deux catégories : les effets marchands (par exemple l'impact sur les fournisseurs ou sur les sous-traitants des entreprises aidées) et les effets non-marchands (par exemple l'impact positif d'une amélioration de l'image

de la région ou l'impact négatif d'une détérioration de l'environnement). Les effets non-marchands, ou externalités ne sont pas pris en compte dans les décisions privées, même s'ils ont des conséquences économiques pour les acteurs privés. Il est donc particulièrement utile de les identifier dans le contexte d'un programme public.

En raison de leur décalage dans le temps ou de leur caractère indirect, les impacts ne peuvent pas être facilement connus des opérateurs dans le cadre de leur gestion quotidienne. Les impacts sont donc principalement analysés à l'occasion des travaux d'évaluation.

Une illustration des termes ressources, réalisations, résultats et impacts est proposée dans l'encadré ci-dessous.

Encadré 24 - Ressources, réalisations, résultats et impacts d'une route

Si on prend l'exemple l'agrandissement d'une route pour le désenclavement d'une vallée, les différentes notions peuvent être illustrées de la manière suivante :

- **Ressources :** Les budgets alloués, les moyens juridiques utilisés pour l'expropriation des propriétaires fonciers, les entreprises mobilisées pour les travaux, les instruments d'information utilisés, etc.
- **Réalisations :** Les kilomètres de voies nouvelles, les carrefours agrandis, les revêtements routiers améliorés, etc.
- **Résultats :** La réduction du temps de trajet pour les véhicules qui voyagent entre deux villes données, l'accroissement de sécurité pour ces véhicules.
- **Impacts :** Le nombre de voyageurs habitant la région et empruntant la route un an après sa mise en service. La meilleure attractivité de la région pour les investisseurs (impact recherché positif). Les nuisances accrues pour certains riverains (impact négatif non recherché). L'augmentation de la valeur de certains terrains avoisinants (impacts positifs non recherchés), etc.

◀ ▶ L'évaluation examine le système d'indicateurs et en particulier s'il reflète suffisamment bien les objectifs (voir les critères de qualité d'un système d'indicateurs). Si nécessaire, et dans un contexte ex ante ou intermédiaire, l'évaluation fait des recommandations pour améliorer les indicateurs.

● Informations relatives au contexte

Les indicateurs de programme concernent tous les aspects de la mise en œuvre et permettent également de suivre les effets (résultats et impacts) depuis les plus directs jusqu'aux plus lointains. Ils vont par exemple mesurer l'atteinte d'un public prioritaire ou l'obtention d'un avantage durable pour les destinataires ultimes du programme. Dans le cadre du suivi et de l'évaluation, un indicateur de programme peut montrer que telle intervention est un succès ou que telle autre est un échec.

A l'inverse, les indicateurs de contexte s'appliquent à l'ensemble de la population sans se préoccuper de savoir si elle a été touchée par le programme ou non. Ils vont par exemple montrer l'ampleur des besoins de la population ou l'importance des disparités à réduire. Ils peuvent servir à apprécier si les objectifs sont suffisamment ambitieux par rapport aux problèmes rencontrés par la population. Dans le cadre d'une évaluation, un indicateur de contexte peut montrer qu'une intervention a perdu sa raison d'être ou doit être modifiée parce que le problème que l'on voulait résoudre a disparu ou a changé de nature.

Le point de rencontre entre les indicateurs d'impact et les indicateurs de contexte semble évident. Ce serait en suivant ces deux indicateurs que l'on pourrait estimer les effets propres du programme. Cependant cette facilité théorique est contrariée par l'existence d'effets dus aux facteurs concurrents dont il va falloir tenir compte. L'état final attendu se décompose en trois parties : état initial, effet attendu et effet des facteurs concurrents.

Pour chaque objectif, l'évaluation examine s'il est associé à un indicateur, s'il existe une quantification de l'état initial, de l'état final attendu, et/ou une quantification de l'effet attendu. Si nécessaire, l'évaluation fait des propositions pour les objectifs qui sont dépourvus d'indicateurs : ou bien, elle propose des indicateurs pertinents ou bien elle propose un traitement qualitatif (voir le § suivant et le Volume 2 de la Collection).

● Quantification des réalisations et des résultats

◀ Un système d'information doit être mobilisé ex ante dans le cadre du processus de négociation du programme afin de préciser les critères de réussite et de quantifier les effets recherchés. Par la suite, le dispositif mis en place pour le suivi doit permettre l'examen des réalisations et des résultats au fur et à mesure qu'ils se concrétisent sur le terrain. Chaque opérateur fournit ses informations aux gestionnaires du programme sous forme de rapports d'avancement, de bilans périodiques ou de comptes-

rendus d'exécution. Les autorités responsables du programme reçoivent ces informations, les discutent dans le cadre du Comité de suivi et réagissent en conséquence.

► Au moment de l'évaluation à mi-parcours et de l'évaluation ex post, il s'agira d'utiliser au mieux les données du système sous réserve de s'être assuré de la qualité des informations fournies par ce système.

L'évaluation va devoir vérifier si des dispositions suffisantes ont été prises qui permettent de vérifier si le programme, tout au long de son cycle de vie, est en train d'atteindre ses objectifs, si ses objectifs restent pertinents et si des décisions seront prises en cas de dérive.

L'évaluation doit également s'assurer que des dispositions suffisantes (de suivi - évaluation) ont été prises pour rendre compte de l'atteinte des objectifs. En effet, ce travail de suivi - évaluation doit avoir été organisé selon un plan conçu avant le démarrage du programme. Il doit être clairement précisé qui est chargé de cette surveillance et de ces décisions et selon quel calendrier. Des ressources financières et humaines auront été prévues à cet effet en quantité suffisante.

C'est grâce à ce cadre que pourra être surveillé le rythme des dépenses et l'avancement du programme (système de suivi) et que pourront être programmées les différentes évaluations susceptibles d'intervenir en cours de programme ou en fin de programme. C'est également ce cadre qui doit préciser comment les évaluations doivent être discutées par les autorités gestionnaires et comment s'organisera leur diffusion auprès du public.

L'évaluation examine les dispositions qui ont été prises pour la quantification des indicateurs. La collecte et la synthèse de l'information peuvent être facilitées par la mise en place d'un système informatisé. Les nouvelles technologies de l'information facilitent la collecte et la circulation d'une grande masse d'informations. Cependant, il faut éviter la construction de systèmes d'indicateurs pléthoriques. L'évaluation peut comporter une réflexion approfondie pour sélectionner les indicateurs les plus utiles.

L'équipe d'évaluation examine la probabilité qu'ont les indicateurs d'être correctement renseignés. Elle fait des propositions pour que, le cas échéant, les indicateurs de ressources et de réalisations soient systématiquement quantifiés par les opérateurs et les gestionnaires dans le cadre de leur responsabilité.

Un certain nombre de critères permettent de vérifier la qualité d'un système d'indicateurs qui sont présentés de façon détaillée dans le deuxième

Volume de la Collection. Pour en résumer brièvement le contenu, on peut préciser qu'il existe un groupe de critères de qualité applicables à chacun des indicateurs retenus d'une part et, d'autre part, un autre groupe qui concerne l'ensemble du système.

Un indicateur est de qualité s'il est quantifié à intervalles réguliers et selon une fréquence appropriée aux travaux considérés et publiés. De plus, l'indicateur doit être utilisable peu de temps après la collecte des données qui en a permis la construction. On doit également vérifier la sensibilité et la fiabilité de ces indicateurs par rapport aux phénomènes que l'on veut mesurer. Enfin, l'indicateur doit permettre de faire des comparaisons entre les mesures du programme mais également avec d'autres programmes ou groupes de destinataires.

Les principaux critères de qualité d'un système d'indicateurs sont au nombre de quatre : bonne couverture de l'ensemble du programme ; bon équilibre entre réalisations, résultats et impacts ; pragmatisme en terme de nombre ; et enfin, pertinence à savoir une sélection d'indicateurs qui tienne compte de l'importance stratégique et opérationnelle des informations.

● Informations qualitatives

Une bonne évaluation ne peut pas se permettre d'exclure de son champ d'analyse un objectif parce qu'il serait difficilement mesurable de façon quantitative. En effet, il se peut que ce soit un des objectifs les plus importants qui ne peut être négligé pour la simple raison que l'on ne sait pas définir un indicateur quantitatif.

Une information qualitative bien structurée est bien évidemment préférable à l'absence d'information. Il convient par conséquent de réaliser un travail de structuration qui peut s'effectuer grâce à une grille d'appréciation plus ou moins structurée si possible graduée selon la plus ou moins grande réussite de l'intervention considérée.

La construction et l'utilisation de ces grilles sont présentées dans le Volume 4 de la Collection MEANS consacré aux solutions techniques adaptées aux programmes socio-économiques complexes. L'encadré ci-dessous donne un exemple d'une grille d'appréciation établie pour la recherche d'impacts environnementaux d'une mesure particulière d'un programme d'Objectif 1 au Royaume-Uni. L'intérêt de cet exemple est l'association d'objectifs économiques avec les autres objectifs.

Encadré 25 – Exemple de grille d'appréciation qualitative de l'impact environnemental des projets

Priorité 1 / Mesure 1 : Sites et locaux de qualité pour investisseurs industriels et commerciaux

Objectif : Garantir la fourniture adaptée de sites industriels et commerciaux de qualité pour encourager la croissance locale et l'augmentation des opportunités d'emploi.

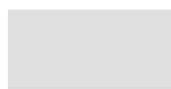
Projet :

Référence :

	Oui	Pas applicable	Non
1. Si le projet est réalisé dans un site industriel réhabilité, est-ce que la zone en friche sera réduite et est-ce que certains bâtiments seront reconvertis en sites de production ?	1	0,5	0
2. Le projet respecte-t-il les consignes relatives à l'environnement naturel ou au patrimoine immobilier ?	1	0,5	0
3. Le projet intègre-t-il de bonnes pratiques d'économie d'énergie, de recyclage de l'eau ou de réduction des déchets, ou implique-t-il de mesures environnementales innovante ?	1	0,5	0
4. Le site bénéficiera-t-il d'un plan de gestion de l'environnement ou bien des mesures sont-elles en place pour la gestion de l'environnement ?	1	0,5	0
5. Des mesures ont-elles été prises pour garantir que le site soit facilement accessible par les transports en commun ?	1	0,5	0
6. Le développement est-il conçu pour améliorer l'environnement immobilier local, par exemple en minimisant les effets de la délinquance ou du manque de civisme ? (ce critère est considéré comme moins important que les autres d'où une note < 1)	0,75	0,38	0
7. Le projet de développement comprend-il des propositions d'améliorations environnementales telles que l'aménagement paysager ? (même remarque que ci-dessus)	0,75	0,38	0
Note totale			

Dans les trois cas, (grille semi-structurée, structurée ou graduée) l'information qualitative peut être transformée sous la forme de notes qui permettent des comparaisons au sein du programme.

◀ ▶ Pour les objectifs qui ne peuvent pas être associés à un indicateur de façon pertinente, l'évaluation peut faire des propositions de traitement qualitatif dans le cas des évaluations ex ante et à mi parcours.



Evaluer les impacts de l'ensemble du programme

Les programmes de développement socio-économiques interviennent habituellement dans des domaines multiples. Bien que leurs impacts convergent vers un objectif global de développement, ils se matérialisent par des mécanismes très différents. Lorsque l'évaluation traite l'ensemble d'un programme, il lui est demandé d'étudier systématiquement tous les impacts. Compte tenu de leur diversité et du nombre des causes possibles, les méthodes applicables sont nécessairement globales.

Ces méthodes reposent sur l'emploi de données secondaires, sur des enquêtes qualitatives auprès des opérateurs et de quelques destinataires, ainsi que sur des extrapolations effectuées à partir de résultats issus d'autres évaluations.

Les programmes financés par la politique européenne de cohésion visent à accélérer le développement socio-économique dans telle ou telle de ses dimensions (revenu, emploi, qualité de vie, etc.). Le principal impact recherché est un impact structurel donc durable au niveau macroéconomique ou macro-social. Pour obtenir cet impact, la plupart des programmes adoptent des stratégies multisectorielles, comprenant des mesures diversifiées autant par leur nature (équipement, conseil, subvention, formation, communication, etc.) que par leur domaine (transports, recherche, tourisme, etc.). Chaque mesure produit d'abord des résultats sur ses destinataires directs, puis ses impacts se matérialisent par étapes successives sur des publics de plus en plus lointains, à travers des mécanismes multiples et ramifiés. L'existence même du programme présuppose que tous ces mécanismes convergent vers l'impact global de développement.

La réalisation d'une évaluation de programme pourra prendre des formes très différentes selon la façon dont on veut analyser ses impacts. Dans cet ouvrage, on a choisi de présenter trois situations comme suit.

1. L'évaluation d'ensemble qui se donne pour contrainte d'examiner les impacts de toutes les mesures. Cette situation est la première qui est décrite ci-après. Sa réalisation est décrite de façon complète, depuis la préparation du travail jusqu'à la rédaction du rapport en passant par la collecte et l'analyse des données.

2. L'approfondissement de l'analyse des impacts macroéconomiques. Cette situation est décrite en second lieu en se limitant uniquement aux spécificités de ce type d'analyse (voir page 145).
3. L'approfondissement de l'analyse d'un impact particulier au niveau microéconomique ou micro-social. Cette troisième situation est également décrite sous forme de complément à une évaluation d'ensemble, plutôt que sous forme d'une évaluation à part entière (voir page 157).

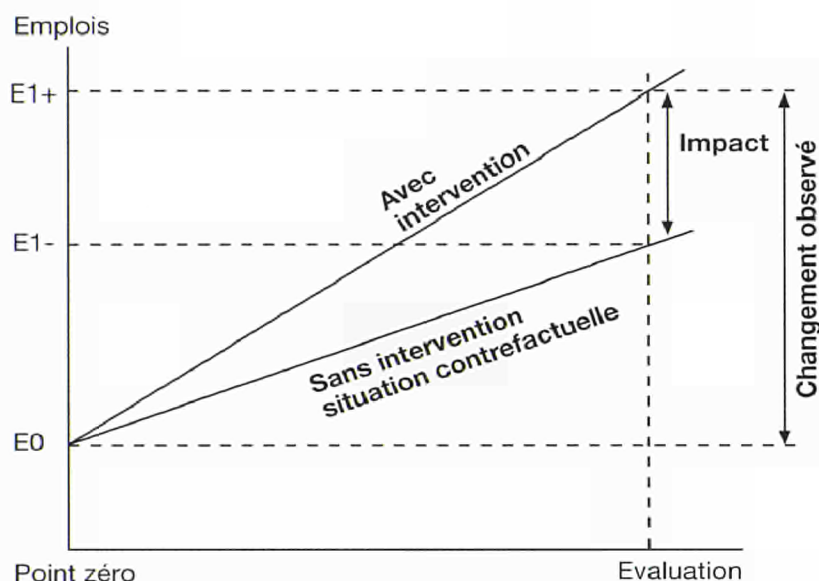
Il est à noter que les deux premières situations (évaluation d'ensemble et analyse macroéconomique) se rencontrent indifféremment à tous les stades du cycle de programmation. Par contre, la situation d'analyse plus approfondie d'un impact particulier est seulement possible de façon rétrospective, à une date suffisamment tardive pour que l'impact ait eu le temps de se matérialiser. Cette analyse est lourde et ne peut pas s'appliquer à l'ensemble d'un programme multisectoriel, comme cela va être expliqué dans les sections suivantes.

■ Choisir le mode d'analyse d'impact

● Imputation d'un changement au programme

Lorsque l'on désire faire l'analyse d'un impact, on ne connaît généralement que ce qui a été observé, c'est à dire un changement (diminution du chômage, nombre d'entreprises créées, augmentation du trafic sur une route, etc.). L'impact ne représente qu'une part de ce changement, c'est à dire la part qui est attribuable au programme et seulement cette part. L'impact est aussi appelé effet net, effet propre ou effet additionnel.

Toutes les méthodes qui permettent d'estimer un impact de façon quantitative consistent, d'une manière ou d'une autre, à décrire ce qui se serait passé en l'absence du programme, à faire une comparaison avec ce qui s'est réellement passé, et à estimer l'impact par différence. En d'autres termes, l'impact du programme est considéré comme l'écart entre l'évolution réelle constatée à la suite de l'intervention (*policy-on*) et l'évolution fictive estimée dans une situation sans intervention (*policy-off*). La situation fictive est aussi appelée « situation contrefactuelle ». Ce raisonnement est illustré dans l'encadré ci-dessous en prenant l'exemple d'un impact sur la création d'emploi.

Encadré 26 - Situation observée et situation contrefactuelle

Une technique d'imputation consiste à créer matériellement les deux situations en recrutant des personnes candidates pour participer au programme et en tirant au sort parmi elles un premier groupe qui participera au programme et un second groupe qui ne participera pas (groupe de contrôle). Une enquête est alors effectuée sur chacun des deux groupes avant et après. L'écart entre les résultats des deux enquêtes donne une estimation fiable de l'impact à condition que les deux groupes soient suffisamment représentatifs (chacun comprendra entre 300 et 1000 individus selon le taux d'erreur acceptable et la plus ou moins grande hétérogénéité de la population enquêtée). Cette technique est longue, coûteuse, techniquement délicate et quelquefois contestable d'un point de vue éthique. Ces raisons expliquent pourquoi elle est rarement appliquée malgré sa très forte capacité à produire des conclusions robustes. A la place, on utilise souvent des techniques qui reconstituent la situation contrefactuelle, non plus de façon réelle, mais de façon artificiellement, comme la technique du groupe de comparaison ou celle de la simulation économétrique qui seront développées plus loin (voir 155 et 166).

L'analyse d'impact est nettement simplifiée si elle est faite directement par les destinataires. C'est ce qui arrive, par exemple, quand on demande à des entrepreneurs qui ont reçu une aide : Quelle sera la valeur ajoutée générée par l'aide dans votre entreprise ? Combien d'emplois ont-ils été créés grâce à l'aide ? Ces questions peuvent être posées au moment de la sélection des projets, au moment du paiement de l'aide, ou encore au moment d'une enquête d'évaluation. Les réponses fournissent une estimation directe des effets et évitent la construction d'une situation contrefactuelle. En effet, c'est la personne enquêtée qui imagine par elle-même la situation contrefactuelle au moment où elle donne sa réponse.

Lorsque l'impact est estimé par cette méthode, il faut toutefois réaliser un supplément d'analyse car les destinataires ont tendance à surestimer les effets du programme pour toutes sortes de raisons, dont la moindre n'est pas la volonté de manifester leur reconnaissance vis à vis de l'administration qui leur a alloué une aide. Dans ce cas, l'équipe d'évaluation procède à une analyse des impacts réels de l'intervention évaluée, c'est à dire de ses « effets nets », par opposition à ses effets apparents ou aux effets qui sont déclarés par les destinataires, c'est à dire les « effets bruts ». Il convient en effet de déterminer dans quelle mesure l'aide attribuée a véritablement permis la réalisation d'un projet, et dans quelle mesure le projet n'aurait pas pu se faire aussi bien et aussi vite en l'absence de l'intervention publique. Lorsque l'aide publique est accordée à des projets qui se seraient réalisés même sans aide on parlera d'un « effet d'aubaine ». pour ceux qui en ont bénéficiée. On obtiendra l'effet net en déduisant l'effet d'aubaine de l'effet brut.

Exemple : L'effet d'aubaine dans le cadre des aides à l'embauche

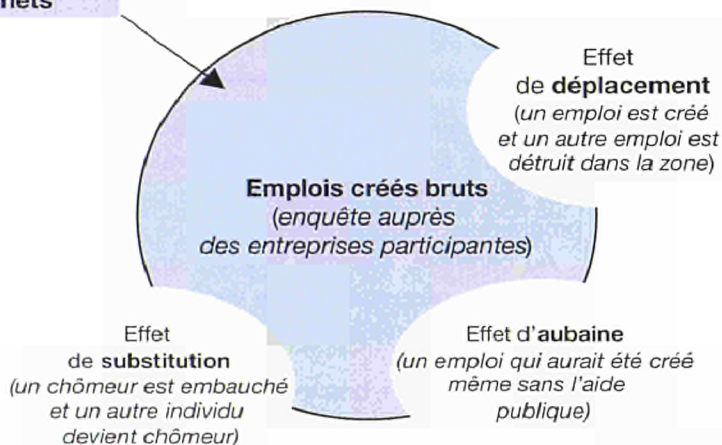
L'exemple le plus souvent connu est celui de l'entreprise qui décide d'une embauche et recherche dans un deuxième temps si cette embauche ne lui permet pas d'obtenir une aide publique à un titre ou un autre. Les nombreuses évaluations qui existent dans le domaine de l'emploi estiment que l'effet d'aubaine pour les aides à l'embauche est très important (entre 40 et 80% des effets bruts). Néanmoins les autorités publiques d'une région ou d'un pays peuvent en toute connaissance de cause décider que le problème social du chômage est prioritaire et que l'effet d'aubaine est un prix qu'il faut accepter de payer pour susciter l'embauche de chômeurs.

Dans certains cas d'évaluation, il est nécessaire d'analyser également les effets de déplacement ou les effets de substitution. L'effet de déplacement apparaît par exemple quand un emploi est créé dans la zone géographique aidée au détriment d'un autre emploi qui a été supprimé ailleurs (Cf. l'émotion

suscitée récemment par la fermeture d'une usine automobile en Belgique et la création simultanée d'emplois en Espagne par le même constructeur). L'effet de substitution apparaît par exemple lorsqu'une personne trouve un emploi grâce à un programme, mais que cet emploi échappe, par le fait même, à une autre personne qui se trouve désavantagée. Les effets de substitution et les effets de déplacement sont quelquefois recherchés, par exemple lorsqu'on encourage le transfert d'entreprises vers des zones qui connaissent un fort taux de chômage. Dans ce cas, l'impact du programme comprend bien évidemment ces effets. En revanche les effets de substitution ou de déplacement doivent être soustraits des effets bruts lorsqu'ils ne sont pas justifiables, par exemple si l'aide apportée à un chômeur aboutit à faire perdre son emploi à quelqu'un qui devient chômeur à son tour. En résumé, et comme le montre l'Encadré 27, l'effet net peut être estimé à partir de l'effet brut à condition de soustraire les effets d'aubaine et, dans certains cas, les effets de déplacement et/ou de substitution si ceux-ci ne sont pas recherchés par le programme.

Encadré 27 - Effet d'aubaine, de déplacement et de substitution

Emplois créés nets



La plupart des effets de déplacement et de substitution peuvent être analysés de façon crédible à l'aide d'entretiens avec les destinataires. Par contre, il est plus délicat d'analyser les effets d'aubaine par cette technique

dans la mesure où il est difficile d'admettre, même sous couvert de l'anonymat, qu'une aide a été reçue sans réelle contrepartie. Une bonne analyse des effets d'aubaine suppose d'utiliser la technique des groupes de comparaison, comme le suggère l'exemple ci dessous.

Exemple : La mise en évidence les effets d'aubaine à l'aide d'un groupe de comparaison

Dans le cadre d'un programme de développement rural mis en œuvre au Danemark, les exploitants agricoles étaient susceptibles de recevoir une subvention pour diversifier leurs activités. Une enquête réalisée dans le cadre de l'évaluation a posé aux exploitants aidés la question suivante : « Etes vous d'accord avec la phrase suivante ? : l'aide reçue a conditionné la réalisation de mon projet de diversification ». Environ 75% des exploitants aidés ont donné une réponse positive à cette question. On peut donc, dans un premier temps, estimer l'effet d'aubaine à environ 25%.

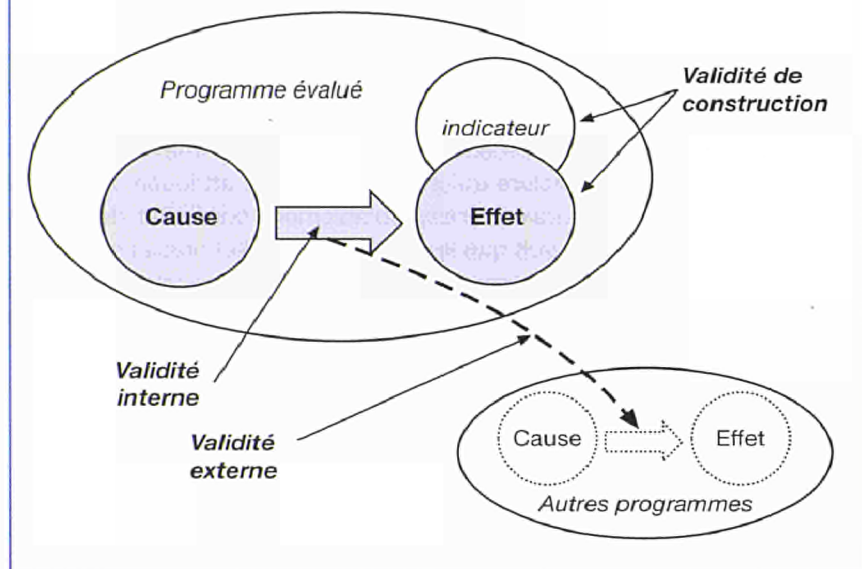
Cependant, un complément d'enquête a été effectué auprès d'exploitants qui avaient demandé l'aide et à qui elle avait été refusée pour diverses raisons. Or il est apparu que tous ces exploitants avaient tout de même réalisé leur projet. On peut donc en conclure que l'effet d'aubaine serait de 100%. Cette conclusion est toutefois hâtive dans la mesure où certains de ces exploitants ont pu être écartés parce que le comité de sélection estimait qu'ils n'avaient pas réellement besoin d'une aide publique. Dans ce cas, un approfondissement de l'analyse, mettant en œuvre un véritable groupe de comparaison, serait nécessaire pour comprendre la réalité des effets d'aubaine.

Ce qui précède montre que l'analyse d'un impact doit être faite avec beaucoup de précautions si on veut qu'elle soit valide. Que l'impact soit estimé à partir des changements constatés ou à partir de l'opinion des participants au programme, l'équipe d'évaluation ne doit jamais s'en tenir aux apparences. Au contraire, elle doit toujours approfondir l'analyse et procéder, chaque fois que nécessaire à une vérification sur le terrain. Une vérification particulièrement crédible est celle qui s'appuie sur une enquête auprès de non-participants.

● **Validité de l'analyse**

Les principaux auteurs de la littérature consacrée à l'évaluation considèrent que la validité d'une analyse d'impact prend trois formes, décrites dans l'encadré ci-dessous : la validité de construction, la validité interne et la validité externe.

Encadré 28 – Les formes de validité des évaluations



La validité de construction porte sur le fait que les indicateurs utilisés pour l'analyse mesurent de façon fiable et sensible les changements que l'on voulait observer. Pour obtenir une bonne validité de construction, il est particulièrement nécessaire que l'impact soit exprimé dans des termes suffisamment précis et que les observations réalisées sur le terrain soient strictement en concordance avec ce que l'on cherche à mesurer. Par exemple, si l'on cherche à estimer l'impact d'une aide sur l'innovation grâce à une enquête auprès des PME régionales, il faut que la notion d'innovation ait été définie de façon précise et que les questions posées correspondent précisément à ces définitions (ex : nombre de nouveaux produits ou procédés de production).

La validité interne est obtenue quand l'équipe d'évaluation apporte une conclusion indiscutable sur le fait que le programme a produit un effet donné. Par exemple, les aides du programme ont entraîné, dans les PME aidées directement et indirectement, un doublement du taux d'innovation (ex : nombre d'innovations par an pour 1 000 emplois). Ce type de conclusion est plus crédible si la collecte et l'analyse des données ont tenu compte très précisément de toutes les particularités du programme et de

son contexte (ex : catégories d'entreprises aidées, formes d'innovations propres à la région, etc.).

On voit par là que pour améliorer la validité interne, il faut adapter l'évaluation à tous les particularismes du programme et de son contexte, ce qui va à l'encontre de la troisième forme de validité, la validité externe, qui consiste à produire des conclusions généralisables à d'autres programmes. Pour poursuivre, l'exemple précédent, l'évaluation aurait une bonne validité externe si elle pouvait conclure qu'une aide similaire, attribuée à des PME dans d'autres régions, auraient vraisemblablement pour effet de doubler leur taux d'innovation. On voit que la validité externe est précieuse dans un contexte où de nombreux programmes appliquent des recettes similaires. Elle permet d'étendre les leçons apprises dans une région et les rendant utilisables dans de nombreuses autres régions.

● Niveaux d'approfondissement

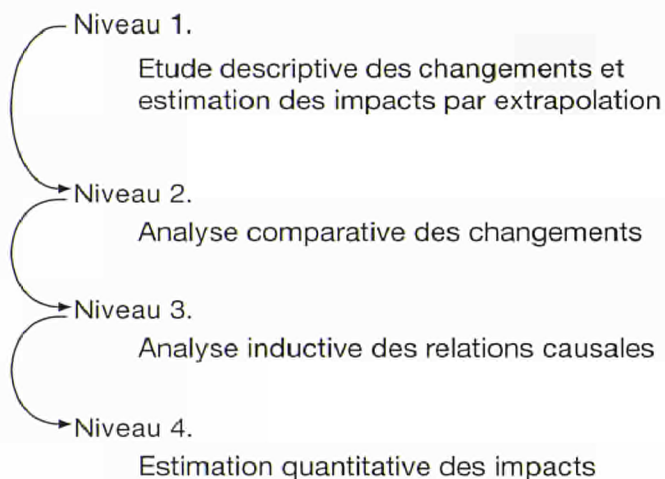
Les sections précédentes ont montré que l'équipe d'évaluation ne peut pas réaliser une estimation d'impact valide sans procéder à des vérifications sur le terrain, à des enquêtes auprès de non-participants et à une analyse fine des modalités de mise en œuvre du programme et de son contexte. Il en découle que l'analyse valide d'un impact demande du temps et un budget suffisant, ce qui peut conduire à l'impossibilité d'analyser tous les impacts d'un programme multisectoriel et multiobjectifs, ou du moins à l'impossibilité de les analyser tous de façon approfondie.

Le choix d'un niveau d'approfondissement est donc particulièrement déterminant. L'évaluation peut être réalisée en se contentant d'études superficielles (niveau 1 de l'Encadré 29 ci-après) ou, au contraire, en mobilisant des moyens lourds pour réaliser des estimations quantitatives (niveau 4). Dans ce cas, on est généralement obligé de concentrer l'analyse sur quelques impacts.

Le traitement le plus simple d'une question évaluative (niveau 1) consiste en une simple description des changements observables auprès des publics concernés ou sur la zone éligible. Il n'est pas rare que l'évaluation se contente d'un tel traitement purement descriptif, alors que l'on sait bien qu'il peut donner une idée fausse des effets du programme. L'analyse peut être améliorée en apportant des correctifs aux données observées de façon à se rapprocher de la réalité des effets propres au programme. Par exemple, on tiendra compte des effets d'aubaine en appliquant un

coefficient correcteur tiré de l'évaluation d'un autre programme, sous réserve d'avoir vérifié la validité externe de cette évaluation. De même, on pourra estimer les impacts sur l'emploi à partir des investissements réalisés par les entreprises aidées, à l'aide d'un coefficient d'extrapolation.

Encadré 29 - Choisir le niveau d'approfondissement de l'analyse d'impact



L'analyse comparative (niveau 2) permet d'aller plus loin dans l'analyse. Elle doit au minimum s'appuyer sur l'observation de publics ou de territoires non touchés par le programme. Les analyses de ce niveau sont toujours possibles, même avec des moyens limités, bien qu'elles ne soient malheureusement pratiquées que dans un nombre encore trop rare de cas. Il est vrai qu'elles ont un coût en termes d'enquêtes de terrain et qu'elles nécessitent un délai d'étude suffisant. Sans démontrer, au sens fort du terme, l'existence d'une relation de causalité entre le programme et les changements socio-économiques intervenus, l'analyse comparative rassemble un faisceau d'indices qualitatifs et quantitatifs sur la nature et l'importance des effets du programme.

L'analyse inductive des relations de causalité (niveau 3) est à considérer lorsque l'on cherche à comprendre les mécanismes de production des impacts et lorsqu'existent de nombreuses interdépendances avec le contexte socio-économique (causes concurrentes). Les analyses de ce

niveau sont toujours possibles, notamment à l'aide d'études de cas, mais elles sont exigeantes en termes d'enquêtes de terrain, de techniques d'évaluation et de délai d'étude.

L'analyse la plus approfondie (niveau 4) est celle qui permet de montrer l'importance quantitative des impacts du programme à partir d'hypothèses sur l'ensemble des relations de cause à effets concurrentes. Cette analyse, lourde et techniquement difficile, nécessite généralement plusieurs mois de travail. De plus, son aboutissement n'est pas toujours assuré pour des raisons expliquées plus loin (voir p 168).

Pour conclure, l'évaluation d'ensemble d'un programme porte systématiquement sur tous les impacts de toutes les composantes d'un programme. Or dans la mesure où les programmes socio-économiques se caractérisent par la multiplicité de leurs impacts, il en découle que leur évaluation d'ensemble ne peut pas espérer dépasser le premier niveau d'analyse. La validité d'une évaluation d'ensemble dépend donc de la validité externe des autres évaluations sur lesquelles elle se fonde pour estimer les impacts, d'une part, et de la réalisation d'analyses plus approfondies lorsque cela s'impose, d'autre part.

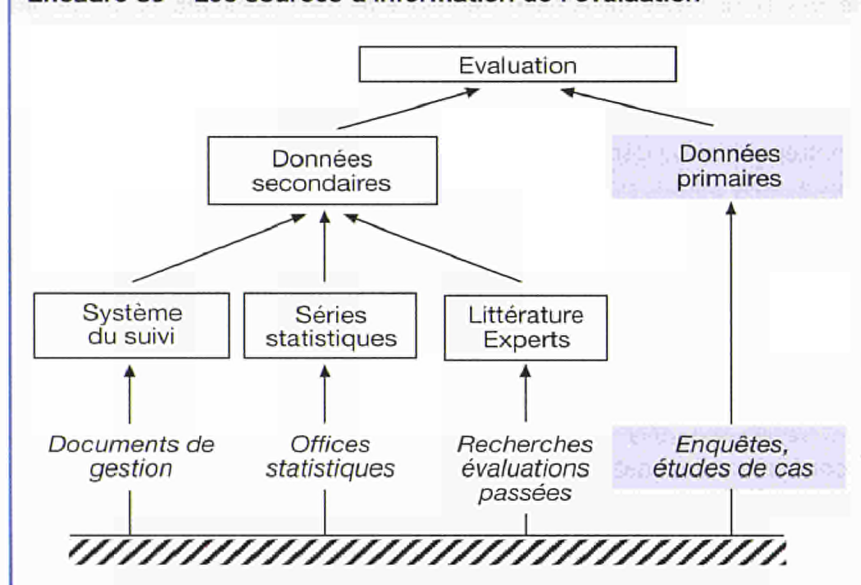
Les analyses plus approfondies (niveaux 2 à 4) permettent d'atteindre un bon niveau de validité, mais leur coût les réserve à l'estimation de quelques impacts. Il est alors d'autant plus important que ces évaluations soient réalisées avec un souci de validité externe, afin que leurs conclusions soient utilisables dans d'autres évaluations, que ce soit dans la même région ou dans des régions connaissant des problèmes similaires.

■ Exploiter les sources d'informations existantes

L'estimation d'un impact est toujours une tâche difficile. Pour aboutir à une estimation aussi robuste que possible, il convient de s'appuyer le plus possible sur les informations existantes telles que les données administratives, les études et les recherches relatives au domaine considéré, la documentation liée à la vie et à la gestion du programme, les statistiques publiées, des informations relatives à des programmes similaires, etc. Ces informations peuvent être obtenues avec peu de nouvelles dépenses et rapidement. Avec un coût de collecte très faible, elles peuvent contribuer à fournir des premières réponses à une partie des questions posées. Si le cahier des charges a été correctement préparé, il contient déjà une liste des informations immédiatement disponibles,

appelées données secondaires par opposition à celles qui vont être collectées spécialement dans le cadre de l'évaluation et qui sont appelées données primaires. Ces deux catégories de données sont présentées dans l'encadré ci-dessous.

Encadré 30 – Les sources d'information de l'évaluation



● Documents de gestion du programme

Les documents de gestion du programme contiennent une information sur les dépenses qui ont été engagées et payées. Ces informations sont utilisables sous forme synthétique, par exemple sous forme de bilans financiers décomposés selon les axes, mesures et actions. Les données élémentaires projet par projet sont systématiquement conservées et accessibles. Il pourra être nécessaire d'y revenir, par exemple dans les cas suivants : (1) on désire connaître le pourcentage des dépenses du programme qui a été consacré à porter remède à des problèmes d'environnement, (2) on désire connaître le pourcentage des dépenses du programme qui a bénéficié aux zones rurales ou urbaines les plus défavorisées, (3) on désire connaître le pourcentage des dépenses du programme qui a bénéficié aux plus petites entreprises, (4) on désire

connaître la différence entre le coût par emploi créé en zone rurale et en zone urbaine, etc.

Parmi les données conservées sur les dépenses, on trouve souvent le nom, l'adresse et le numéro de téléphone des personnes et des organisations qui ont reçu les aides. Ces données sont très précieuses pour l'évaluation dans le cas où il serait nécessaire de contacter des destinataires à des fins d'enquête.

Les documents de gestion du programme contiennent également une information sur les réalisations, c'est à dire sur ce qui a été obtenu en contrepartie de la dépense publique. Ces informations sont utilisables sous forme synthétique, par exemple, à travers des indicateurs ou des rapports d'activité. Les données élémentaires, projet par projet, sont souvent conservées mais généralement de manière peu accessible. Il pourra cependant être nécessaire d'y revenir, par exemple si l'on désire connaître le détail de quelques projets pour réaliser des études de cas.

L'équipe d'évaluation ne peut accéder aux documents de gestion qu'avec l'aide et la coopération de ceux qui les détiennent, ce qui peut représenter un temps de travail non négligeable pour les gestionnaires et les opérateurs. La coopération des détenteurs de l'information est obtenue d'autant plus facilement que le problème a été discuté à l'avance avec les intéressés. L'autorité commanditaire est responsable de faire en sorte que les portes s'ouvrent, ce qui peut être obtenu par exemple en associant les principaux détenteurs d'information au groupe de pilotage ou en promettant un retour sous forme de synthèse des premières conclusions de l'évaluation. Pour faciliter l'accès à l'information, il est également important que l'équipe d'évaluation s'engage à préserver la confidentialité de toutes les données personnelles.

● Sources statistiques

Les sources statistiques concernent la totalité d'une région, d'un Etat, voire de l'Union européenne. Elles permettent donc d'examiner des changements qui affectent l'ensemble d'une population, d'une économie, de l'environnement d'un territoire donné ou d'une population particulière au sein d'un territoire. Les statistiques les plus utiles à l'évaluation sont celles qui s'appliquent exactement au territoire concerné par le programme ou à la population éligible. Ces dernières sources statistiques fournissent des indicateurs de contexte et permettent de constater des impacts apparents (par exemple : le nombre de créations d'entreprises a doublé) ou

d'apprécier des besoins (par exemple, le taux de création d'entreprises est nettement inférieur à la moyenne européenne), ou encore d'examiner si les objectifs restent pertinents (par exemple, le taux de création d'entreprises a maintenant rattrapé la moyenne européenne).

A la différence des données de gestion qui se concentrent sur les opérateurs et sur les destinataires directs du programme, les statistiques englobent aussi toutes les personnes ou les entreprises qui n'ont eu aucun contact avec le programme. La comparaison d'une statistique avant-après ne peut donc absolument pas fournir une estimation d'impact. Elle donne tout au plus une information utilisable dans le cadre d'une analyse d'impact dont le rôle sera d'imputer le changement observé à plusieurs causes : le programme d'une part, et les facteurs exogènes (ou facteurs concurrents) d'autre part. Ainsi par exemple, les statistiques peuvent montrer une augmentation du chômage explicable par un fort accroissement naturel de la population active alors même que le programme a créé beaucoup d'emplois.

Dans le cadre d'une analyse d'impact, les statistiques fournissent des indications utiles sur l'évolution des facteurs exogènes en mesurant toutes sortes de caractéristiques du territoire ou de la population concernée. Elles permettent d'interpréter ou de nuancer l'observation des effets bruts ou impacts apparents. Elles fournissent aussi des coefficients d'extrapolation dont on verra qu'ils sont souvent utilisés en évaluation (voir page 134). Par exemple, si une étude statistique permet d'affirmer que la taille moyenne des entreprises créées depuis deux ans s'élève à 4,5 emplois, ce coefficient pourra être utilisé dans le cadre d'une estimation d'impact pour une mesure d'aide à la création d'entreprises.

Les données statistiques sont accessibles directement auprès des organismes qui les produisent et les publient (offices statistiques publics aux niveaux européen, national et régional, instituts privés, etc.). Elles sont souvent déjà rassemblées par les gestionnaires du programme, ou par des organismes d'étude (observatoires régionaux par exemple). Par exemple, un grand nombre de données régionalisées peut être trouvé à Eurostat dans la base REGIO (base de données dont l'accès est payant). Ces données sont présentées sous la forme d'environ 70 tableaux standards par domaine : démographie, comptes économiques, chômage, force de travail, énergie, agriculture, transport, R&D. Eurostat a récemment enrichi ses données régionalisées grâce à l'exploitation des enquêtes nationales annuelles sur les structures industrielles. L'information traitée est disponible dans la base de données ELIZE (base de données qui ne dépend pas

d'Eurostat). Celle-ci comprend : le nombre d'entreprises industrielles, l'emploi, les salaires et traitements, le chiffre d'affaires et l'investissement. Les données sont décomposées en une trentaine de secteurs économiques et en deux classes de taille (± 20 employés). A moyen terme, la base de données REGIO sera complétée par une exploitation de l'enquête sur les structures agricoles.

L'utilisation des statistiques par l'équipe d'évaluation peut poser plusieurs sortes de problèmes techniques. En premier lieu, pour des raisons de secret statistique ou de méthode de collecte, on ne peut pas toujours régionaliser finement les statistiques. Ainsi, les enquêtes construites sur des échantillons limités (10 000 à 40 000 questionnaires à l'échelle européenne) ne peuvent donner de résultats significatifs que pour des grandes régions (enquête sur les forces de travail, enquête innovation, Eurobaromètre, pour citer quelques exemples). Même lorsque les statistiques sont finement régionalisées, leur découpage ne coïncide pas toujours avec celui des programmes. Enfin, en l'état actuel de la pratique, certaines données sont calculées et publiées avec deux ou trois ans de retard. Pour des suggestions sur la façon de traiter ces difficultés techniques, on se reportera au Volume 2 de la Collection.

● **Evaluations et recherches passées**

Les informations que l'on peut retrouver dans les évaluations et recherches passées peuvent jouer un grand rôle à toutes les étapes de l'évaluation. Par exemple, un retour à la littérature spécialisée pourra permettre de suggérer un indicateur pertinent, une grille d'observation déjà testée, un modèle explicatif des impacts, un coefficient d'extrapolation, une référence de comparaison, etc.

Il peut s'avérer très utile d'avoir recours aux conclusions d'évaluations des programmes similaires dans la même région ou dans d'autres régions. Les mêmes causes pouvant engendrer les mêmes effets, se référer à des programmes similaires peut souvent permettre de répondre rapidement et à peu de frais à une question évaluative. Si l'équipe d'évaluation s'appuie fortement sur ce type de données secondaires, elle doit alors faire une vérification de la validité externe des évaluations consultées, c'est à dire de la possibilité de généraliser leurs conclusions à d'autres contextes.

Il n'est pas souhaitable d'investir trop de temps pour inventorier et collecter les données bibliographiques exhaustivement. La façon la plus efficace d'accéder à ces informations est de consulter quelques experts ayant une

connaissance approfondie du domaine et/ou de la région concernée par l'évaluation.

■ Collecter des informations nouvelles

Les données préexistantes, ou données secondaires, répondent rarement à la totalité des questions que l'on est amené à se poser dans le cadre d'une évaluation. Dans certains cas, la bonne qualité du système de suivi suffit à évaluer les résultats et l'abondance des travaux de recherche disponibles permet d'extrapoler les impacts. Dans d'autres cas, il faut envisager des investigations spécifiques et une collecte de données primaires par enquête sur le terrain.

● Entretiens avec les gestionnaires et les opérateurs

Les gestionnaires du programme sont les personnes ou les organisations qui contribuent à la mise en œuvre du programme. Ils attribuent et gèrent les fonds. Les opérateurs, situés au plus près terrain, mettent en œuvre le programme. Les destinataires comprennent, de façon large, toutes les personnes ou organisations qui reçoivent des avantages du programme de façon directe ou indirecte. Ces différents types d'acteurs, qui parfois se confondent (voir l'encadré ci-dessous), étant détenteurs d'informations à différents niveaux du programme, ils seront sollicités dans le cadre de la collecte d'informations.

Encadré 31 – Différence entre opérateurs et destinataires

Selon les définitions de cet ouvrage, le promoteur de projet ou porteur de projet peut être opérateur ou destinataire selon les cas. Dans l'exemple où le promoteur de projet est un groupement d'entreprises qui reçoit une aide pour développer des actions collectives de promotion des exportations, deux situations peuvent se présenter. Si l'aide est ponctuelle et limitée à un seul projet, alors le groupement est considéré comme destinataire direct et les entreprises qui le composent comme destinataires indirects. Si le groupement répète ses actions d'année en année, reçoit des aides variables en fonction de ses activités, fait un compte rendu annuel de ses actions ainsi que de l'identité des entreprises participantes, alors le groupement est considéré comme opérateur et les entreprises participantes comme destinataires directes (ex : une Chambre de commerce et d'industrie qui organise des stages à l'intention des PME).

Lors d'une enquête destinée à collecter des données primaires, l'équipe d'évaluation n'abordera pas toujours le groupement d'entreprises de la même façon. Elle ne posera pas les mêmes questions, selon qu'il a un rôle d'opérateur ou de destinataire.

Les données de gestion sont des données secondaires, dont il a été question précédemment, et auxquelles on accède par les questionnaires. Ce qui est évoqué ici est différent. Il s'agit essentiellement de l'opinion des gestionnaires et des opérateurs sur les questions intéressant l'évaluation, opinion qui est collectée sous forme d'entretiens. Parmi les multiples sortes d'opinions que l'évaluation est amenée à collecter, on mentionnera spécialement celles qui portent sur les succès et les échecs rencontrés par le programme.

Les questions que l'équipe d'évaluation pose dans ces entretiens sont du type suivant : Quelles mesures, quelles actions, quels projets peuvent être considérés comme des réussites exemplaires (ou comme des échecs dont il faut tirer les leçons) ? Comment ces réussites et ces échecs s'expliquent-ils ? A quoi reconnaît-on qu'un projet est particulièrement réussi ? Quels sont les critères spontanément utilisés dans ce jugement (efficacité, efficience, pertinence) ?

L'information qualitative collectée à l'occasion de ces entretiens complète utilement les données de gestion, notamment pour identifier des exemples de bonnes pratiques et pour orienter les analyses d'impact vers les situations où il y a le plus à apprendre. A ce stade de son travail, l'équipe d'évaluation doit être très attentive à privilégier l'examen des effets du programme. Les gestionnaires et les opérateurs ont en effet tendance à considérer qu'un projet est un succès ou un échec en fonction des difficultés plus ou moins grandes rencontrées dans l'application des règles ou dans le fonctionnement de la machinerie administrative. Au contraire, l'évaluation juge la réussite d'un projet en fonction de ses résultats bénéfiques et durables pour les destinataires. Les deux formes de réussite sont souvent liées mais il peut arriver qu'un projet soit problématique du point de vue administratif (interprétation des textes à la limite de la régularité, conflits entre gestionnaires) et qu'il soit en même temps une réussite exemplaire du point de vue de son efficacité (L'inverse se constate encore plus fréquemment !).

L'avis des gestionnaires et des opérateurs s'obtient le plus souvent à travers des entretiens individuels. L'équipe d'évaluation s'efforce de conduire suffisamment d'entretiens pour rencontrer au moins une personne fortement concernée par chaque objectif à chaque niveau de décomposition du programme. Il peut être intéressant, dans cet esprit, de reporter visuellement les entretiens effectués sur l'arbre des objectifs. Selon la complexité du programme le nombre d'entretiens varie de quelques-uns à plusieurs dizaines.

Avec l'accord du commanditaire, l'équipe d'évaluation peut compléter ses entretiens individuels par une série de réunions, chacune rassemblant une dizaine de gestionnaires ou d'opérateurs. Ces réunions peuvent être organisées par ville, par sous-région, par secteur économique, par domaine d'intervention, etc. Il peut également être intéressant d'organiser séparément les réunions aux différents niveaux du programme (opérateurs de terrain, gestionnaires des mesures, responsables d'ensemble du programme). Un exemple de cette dernière démarche pourra être trouvé dans le Volume 4 de cette Collection, consacré aux techniques innovantes d'évaluation des programmes complexes. On trouvera également dans le Volume 3, consacré aux outils d'évaluation, plusieurs techniques d'animation ou de médiations susceptibles de rendre ces réunions plus productives et plus impartiales.

● Entretiens avec les destinataires

Par définition, les programmes socio-économiques couvrent simultanément plusieurs domaines d'intervention (plusieurs secteurs) et de nombreux publics, souvent hétérogènes (ex : PME, créateurs d'entreprises, agriculteurs, travailleurs indépendants, salariés menacés de chômage, chômeurs, chercheurs, etc.). Par conséquent, l'évaluation d'ensemble du programme couvre systématiquement tous les domaines et tous les publics. Cela implique qu'il faut trouver un moyen de collecter des informations réellement nouvelles auprès de ces publics très hétérogènes sans pour autant que le coût d'enquête soit exorbitant. Or si l'on choisit par exemple de faire une enquête par questionnaire, construire un échantillon représentatif pour chacun des publics concernés représente un coût extrêmement élevé. Cela ne peut être fait en pratique que dans le cadre de l'approfondissement d'une question évaluative centrée sur un public particulier (voir la partie suivante).

Parmi les techniques envisageables de façon pragmatique pour collecter des données primaires auprès de destinataires très différents les uns des autres, on citera l'enquête par entretien de groupe (*focus group*) et l'enquête par étude de cas, qui est présentée de façon plus détaillée dans les paragraphes suivants.

La technique de l'entretien de groupe a été standardisée par les professionnels de l'étude de marché. Un entretien de groupe réunit six à huit participants pendant une heure et demi à deux heures autour d'un

enquêteur / animateur. Celui-ci provoque une forte interaction entre les participants sur les thèmes centraux de l'étude. En faisant appel à la créativité des participants et à la dynamique de groupe, la technique permet de démultiplier la capacité d'exploration puis de dégager les idées convergentes. Si le groupe est bien encadré, ce qui émane de la discussion est souvent beaucoup plus riche que ce que l'on obtient d'une série d'entretiens individuels.

La présence de plusieurs personnes évite les opinions extrêmes puisque chaque argument est, le plus souvent, contesté. Cependant, le clientélisme des participants, la présence de personnalités « *leader* » ou une situation conflictuelle entre certaines personnes peuvent entacher la qualité d'une réunion. C'est la raison pour laquelle l'entretien de groupe est plus difficile à réaliser avec des gestionnaires et des opérateurs. Pour une enquête auprès des destinataires, le risque est moins important et peut être prévenu en réalisant une vague de trois ou quatre réunions. Pour plus d'informations sur cette technique, on peut se référer au Volume 3 de la Collection.

■ Analyser l'ensemble des impacts sur les destinataires

Il convient de rappeler au début de cette section que nous nous situons dans le cadre de l'évaluation d'ensemble d'un programme qui comprend des réalisations hétérogènes et qui produit des impacts multiples sur des publics variés. Cette situation empêche l'emploi de la plupart des techniques classiques d'analyse d'impact, lesquelles seront présentées plus loin (voir page 164). Par ailleurs, il convient de distinguer les impacts selon deux niveaux, à savoir ceux qui peuvent être suivis à la trace chez les bénéficiaires du programme, d'une part, et ceux qui ne se matérialisent qu'à l'échelle de l'ensemble d'une région, d'autre part. Le premier niveau d'impact relève de mécanismes microéconomiques ou micro-sociaux, tandis que le second niveau relève de mécanismes macroéconomiques ou macro-sociaux. Leur analyse fera l'objet d'une section spéciale ci-après (voir page 145).

● Croisement des données primaires et secondaires

Pour réaliser son analyse d'impact, l'équipe d'évaluation commence par constater les changements qui se sont déroulés pendant le programme et par s'interroger sur leur concomitance avec l'action politique. Il faut être conscient du fait qu'à ce stade de l'évaluation il est difficile de trouver des

preuves concernant des relations de causalité. Aussi, cherche-t-on plutôt des présomptions de causalité et des explications aux phénomènes constatés.

Ainsi par exemple, il est intéressant de croiser les données statistiques et les données de gestion dans une perspective avant/après comme dans l'exemple suivant.

Exemple : un croisement de données qui révèle un impact plausible

Dans le cadre d'une analyse d'évaluabilité portant sur plusieurs mesures du Programme Opérationnel FEDER de l'Objectif 1 (1994-99) de la région de Valencia (Espagne), un certain nombre de critères d'évaluation ont été déterminés par type d'action. Pour chacun des critères, entre trois et onze indicateurs accompagnés de leurs sources d'information ont été également proposés. Le nombre total d'indicateurs proposés s'est élevé à 123. Certains d'entre eux, environ 40, ont été supprimés dès le début de l'évaluation intermédiaire car ils n'étaient pas disponibles.

Une sélection a ensuite été réalisée par les chargés d'évaluation parmi les indicateurs restants. L'équipe d'évaluation a veillé d'une part, à ce que les indicateurs retenus fournissent une information contrastée (région de Valencia et Espagne) et d'autre part, à ce qu'ils soient basés sur des informations publiées de façon périodique (pour l'analyse de la situation tendancielle 1990-1993) et sur des informations non publiées mais disponibles pour la période la plus récente. Les indicateurs ainsi définis figurant dans le rapport d'évaluation ont fait l'objet d'une analyse approfondie à l'aide de graphiques qui ont permis, d'une part d'examiner la tendance (favorable ou défavorable) pour la période précédant la mise en œuvre du programme, et d'autre part d'identifier s'il y a eu une variation de tendance les années suivantes. Le changement de tendance a été classé comme positif lorsqu'il s'agit d'une amélioration de la situation de la région ou négatif lorsqu'il s'agit d'une détérioration. L'intensité de la variation a été qualifiée d'insignifiante, de modérée ou d'intense.

En guise d'exemple : pour le critère « diversification et efficience des ressources énergétiques », un des indicateurs utilisés a été le nombre d'entreprises consommant du gaz naturel (% région de Valencia par rapport à l'Espagne). Pour la période 1990-93, cet indicateur a révélé une tendance défavorable. A partir de 1994, lorsque la mise en œuvre du programme a débuté, une variation positive de tendance vers une intensité modérée est observée.

► On peut réaliser des croisements particulièrement intéressants à l'aide d'un outil qui devient de plus en plus répandu, à savoir le système d'information géographique (SIG). Cette technique combine une base de données informatisée capable de mémoriser aussi bien les données de gestion du programme que des données statistiques de contexte.

L'utilisation d'un SIG permet de produire des cartes qui effectuent des recoupements visuels entre les données. Ainsi, une carte peut superposer l'évolution du chômage dans différentes zones urbaines ou rurales couvertes par le programme et des informations susceptibles d'expliquer cette évolution, y compris l'importance des financements alloués, le nombre de chômeurs formés et le nombre d'entreprises aidées. Pour plus d'informations sur cette technique, on se reportera au Volume 3 de la Collection MEANS.

● Exploration des mécanismes d'impact

L'enquête par étude de cas examine un nombre limité d'actions, de projets ou de situations (les cas) et aboutit à une série de monographies décrivant chacun de ces cas en quelques pages. Pour chacun des projets étudiés, l'enquêteur rassemble toutes les données de gestion disponibles, interroge les gestionnaires et opérateurs responsables, et réalise des entretiens avec plusieurs destinataires directs et indirects. Chaque étude de cas représente une dépense non négligeable mais le nombre de cas étudiés est suffisamment réduit pour contenir le coût total de l'évaluation dans une limite raisonnable.

Une enquête par étude de cas permet d'obtenir des explications sur l'attitude des destinataires, sur leurs véritables besoins, sur les raisons de leur participation au programme, sur leur façon de juger la qualité du programme, sur les mécanismes d'apparition des impacts, sur les effets imprévus, etc. L'enquête par étude de cas étant essentiellement qualitative, sa principale fonction est de mieux faire comprendre le fonctionnement du programme. Elle aide aussi à interpréter des informations quantitatives obtenues par ailleurs.

Dans certaines limites, on peut utiliser l'étude de cas pour produire une information quantitative. De fait, chaque monographie de cas peut produire une estimation quantitative d'impact, évidemment limitée au projet étudié. La question est alors celle de la généralisation des enseignements d'un cas. Elle peut être traitée de différentes façons au moment de la sélection des cas. Par exemple, si chaque cas est sélectionné comme étant « typique » d'une catégorie de projets, son impact sera considéré comme une image ou un ordre de grandeur de l'impact de tous les projets de cette catégorie. Plus intéressante est sans doute la démarche qui consiste à étudier deux cas par catégorie de projets, un cas de réussite et un cas d'échec, de façon à obtenir une fourchette d'estimation des effets du

programme. Pour une présentation plus complète de la technique de l'étude de cas, on se reportera au Volume 3 de la Collection MEANS et pour un exemple d'utilisation quantitative des études de cas, on se reportera au Volume 4.

● Estimation quantitative par extrapolation

L'estimation quantitative d'un impact, ou au moins la démonstration de l'existence de cet impact, constitue un thème central dans la démarche d'évaluation. Les techniques de base employées pour ce type d'analyse sont les groupes de comparaison et l'analyse de régression, ainsi que la modélisation pour les impacts macroéconomiques. Ces techniques seront présentées plus loin. Dans cette section, on privilégie la seule approche qui soit applicable de façon pragmatique à l'évaluation d'ensemble d'un programme complexe, à savoir l'utilisation de coefficients d'extrapolation pour estimer les impacts.

Un coefficient d'extrapolation permettra par exemple d'estimer le nombre d'emplois temporaires créés par million d'Euros d'investissement en infrastructure, la valeur ajoutée annuelle générée par million d'Euros d'investissement dans les PME ou la proportion des créations d'emploi déclarées par les entreprises qui sont des créations réelles. Si l'équipe d'évaluation emploie des coefficients d'extrapolation préexistants, il importe qu'elle explique et justifie ceux qu'elle a choisis. Un coefficient peut être déduit d'une enquête approfondie conduite dans le cadre de la même évaluation ou d'une évaluation précédente. Un coefficient peut aussi être emprunté à une évaluation ou à une recherche portant sur des interventions similaires, à conditions que les conclusions soient considérées comme généralisables. Un coefficient peut également être établi pour les besoins d'une évaluation par un panel d'experts ou par un réseau d'experts en utilisant la technique de l'enquête Delphi (Voir le Volume 3 de la Collection MEANS pour en savoir plus sur cette technique de consultation d'experts). Les différentes solutions peuvent également être combinées entre elles.

◀ L'extrapolation consiste à appliquer un coefficient d'impact à une donnée de départ qui peut être de nature différente selon les cas. Dans une situation d'évaluation ex ante, l'extrapolation ne peut partir que de données basiques sur les enveloppes budgétaires que le programme prévoit d'allouer aux différentes interventions. On utilise donc des coefficients du type « nombre d'emplois créés par million d'euros » (ou le ratio inverse du coût par emploi créé).

► Dans une situation d'évaluation intermédiaire ou ex post, il est possible d'utiliser les données de réalisation comme base d'extrapolation. Ainsi par exemple, on se fondera sur le nombre de lits d'un équipement touristique récemment construit, pour estimer le nombre d'emplois créés à l'aide d'un ratio de nombre d'emplois par lit. Dans certains cas, il est demandé aux destinataires de faire une déclaration sur les impacts au moment où ils sont en contact avec les opérateurs. Ils doivent par exemple indiquer le nombre d'emplois créés au moment du dernier paiement ou le nombre d'emplois à créer au moment de leur demande d'aide. Dans ce cas il est recommandé d'appliquer un coefficient correcteur pour tenir compte des effets d'aubaine. Ce coefficient peut varier de 0,8 dans les meilleurs cas, à moins de 0,3 pour certaines interventions où les effets d'aubaine sont particulièrement importants.

● Analyse des effets de synergie

Un programme est composé d'un ensemble de projets et de mesures qui sont tous orientés vers le même objectif. Ces projets ne sont pas indépendants et ils ne doivent pas être évalués en faisant l'addition, ou même la synthèse, de leurs effets pris séparément. Au contraire, les composantes d'un programme sont censées se renforcer mutuellement grâce à des effets de synergie. L'encadré ci-dessous décrit quelques mécanismes de production de ces effets de synergie.

Encadré 32 – Exemples de mécanismes produisant des effets de synergie au sein d'un programme

Sous le titre de *Program Induced Dynamic Effects*, une étude réalisée pour la Commission européenne propose une liste des mécanismes qui sont susceptibles de produire des effets de synergie :

- effet de masse critique : un projet isolé n'aurait aucun impact faute d'une masse critique suffisante ;
- effet de taille : abaissement des coûts du fait de la présence de plusieurs projets au même endroit ;
- effet de démonstration : démultiplication de l'action par l'exemple ;
- effet d'ajustement : meilleure coordination des différentes actions publiques entre elles ;
- effet de capacité : une action publique isolée n'aurait aucun impact faute d'une masse critique suffisante ;
- effet de productivité : les économies d'échelle sur l'action publique elle-même améliorent sa productivité ;
- effet de déblocage : la réalisation d'un projet débloque une contrainte pesant sur un autre projet ;
- effet d'entraînement : la réalisation d'un projet incite à réaliser un autre projet.

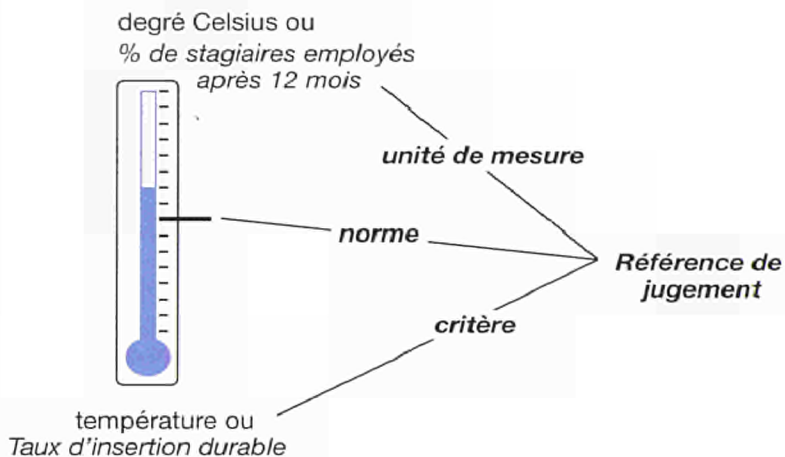
Bien que la synergie soit un mécanisme important qui conditionne l'atteinte des objectifs d'un programme, elle fait rarement l'objet d'une analyse en tant que telle, notamment parce que cela implique un coût supplémentaire alors même que la question ne peut être traitée que dans le cadre d'une évaluation d'ensemble. Pour traiter cette difficulté, un travail méthodologique a été mené dans le cadre du programme MEANS. La démarche proposée est présentée en détail dans le Volume 4 de la Collection. Elle débute par la construction d'une matrice des impacts croisés où toutes les composantes du programme figurent de façon symétrique en lignes et en colonnes. À l'aide d'entretiens avec les gestionnaires et les opérateurs, l'équipe d'évaluation estime le degré d'importance des synergies potentielles entre composantes (sur une échelle de notation de -3 à +3). Ces estimations permettent de remplir les cellules de la matrice et de calculer un coefficient de synergie pour chaque composante. Certaines synergies potentielles peuvent alors faire l'objet de vérifications sur le terrain, particulièrement si elles constituent un enjeu important en termes de compréhension des effets du programme ou en termes de décisions à prendre.

■ Formuler les jugements et conclure l'évaluation

● Références de jugement

Le mot évaluation, apparenté au mot valeur, sous-entend qu'une évaluation est, par définition, un jugement de valeur. Toute la crédibilité de l'évaluation repose sur l'explicitation des références de ce jugement, étape indispensable à toute évaluation et trop souvent négligée. Un jugement de valeur se réfère à un ou plusieurs critères, chacun d'entre eux étant éventuellement associé à un instrument de mesure et à une norme à atteindre, comme le symbolise le thermomètre présenté dans l'Encadré 33.

Chaque évaluation et chaque réponse à une question évaluative peut être réalisée au regard d'un critère d'évaluation spécifique qui est choisi pendant la phase de conception de l'évaluation. Cependant, il y a certains critères d'évaluation qui ont une place prédominante parce qu'ils sont récurrents et essentiels. On les retrouve dans presque toutes les évaluations. Ces principaux critères sont la pertinence du programme et des actions vis à vis des besoins, l'efficacité du programme pour remplir les objectifs et l'efficience du programme, à savoir sa capacité à remplir les objectifs au moindre coût (Voir page 73).

Encadré 33 - Référence, norme et critère

Une fois les critères choisis et une fois leurs normes respectives arrêtées, les instruments qui vont permettre de mesurer sont en règle générale, les indicateurs, et dans certains cas, la notation.

Si l'évaluation du programme pouvait se faire selon un seul critère, par exemple l'efficacité en termes de coût par emploi créé, la formulation du jugement pourrait alors s'appuyer sur une seule technique, en l'occurrence l'analyse coût-efficacité, qui permettrait de juger le programme au regard de ce critère.

Cependant, dans le cadre de l'évaluation de programme où les impacts ne sont pas tous connus de manière très précise, une telle technique, plus appropriée dans les cas d'évaluations approfondies d'actions simples, ne peut pas donner de résultats très certains. Il est alors généralement conseillé de retenir le coût comme critère d'efficacité mais d'utiliser ce critère dans l'optique d'établir des comparaisons. On calculera le coût moyen par emploi créé et on comparera ce ratio avec d'autres ratios obtenus ailleurs dans le cadre de programmes ayant des caractéristiques similaires. On définira le ratio « idéal », en général le meilleur, qui deviendra alors la « norme ». Ce principe de l'étalonnage (*benchmarking*) qui nécessite de rechercher et de valider des références, est également une technique qui permet d'établir un jugement de valeur en référence à un modèle reconnu pour sa qualité. La démultiplication des objectifs oblige à l'identification de multiples cas de bonne pratique, identifiés chacun pour leur qualité dans un domaine particulier.

● Hiérarchisation des critères

L'équipe d'évaluation est souvent confrontée à la nécessité de prendre en compte différents critères et de faire un arbitrage entre ces critères au moment de rédiger ses conclusions. Par exemple, que faut-il conclure à propos d'une mesure qui coûte cher, qui crée peu d'emplois, mais qui touche un public très défavorisé ? Ces arbitrages doivent être faits de façon explicite et justifiée sous peine de mettre en cause l'impartialité de l'évaluation. En cas de difficulté, il vaut mieux que l'équipe d'évaluation demande à être guidée par le groupe de pilotage pour arrêter les critères prioritaires. Si le groupe de pilotage a lui-même des difficultés à arbitrer sur des sujets jugés trop sensibles, il peut faire appel à un médiateur. A cet égard, il est possible d'employer certains outils d'évaluation comme technique de médiation. Des outils comme METAPLAN® ou l'analyse SWOT (forces-faiblesses-opportunités-menaces) ont la caractéristique de faire s'exprimer toutes les opinions et d'identifier sans heurts les points de consensus (Voir Volume 3 de la Collection).

On peut également avoir recours à l'analyse multicritère qui permet de prendre en compte simultanément plusieurs critères en explicitant leurs inter-relations. L'utilisation de cette technique est fréquente pour l'évaluation des projets. Son emploi relève encore de l'innovation pour l'évaluation d'ensemble d'un programme multi-décideurs. Un exemple d'application de cette technique est décrit dans le Volume 4 de cette Collection. Dans le cadre d'une évaluation de programme, l'analyse multicritère implique la participation active d'un groupe de pilotage, si possible élargi à une dizaine de membres. L'analyse est utilisée pour aider les participants à formuler collectivement un jugement comparatif sur les différentes mesures du programme. Outre sa capacité à produire des conclusions consensuelles, la technique joue aussi un rôle utile de clarification et d'explicitation des critères de jugement.

L'analyse coût-avantages propose une autre approche pour hiérarchiser les critères. Chaque impact est pondéré par l'équipe d'évaluation à l'aide d'une valeur monétaire représentant ce que la société ou les personnes concernées sont prêtes à payer pour obtenir les impacts positifs ou éviter les impacts négatifs. L'emploi de cette technique est souvent délicat, à cause notamment de la difficulté de monétariser certains impacts de façon incontestable (impacts sur l'environnement notamment). De plus, la technique est extrêmement difficile à appliquer à un large programme et se trouve pratiquement réservée à l'évaluation des projets d'infrastructures.

Lorsqu'il paraît trop difficile ou trop délicat d'engager l'évaluation dans la voie d'une explicitation et d'une hiérarchisation des critères, une bonne solution alternative consiste à faire formuler les conclusions de l'évaluation par un panel d'experts. Dans des domaines comme ceux des politiques de recherche ou d'éducation, la technique du panel d'experts est largement utilisée. Les experts sont choisis de façon à représenter tous les points de vue en présence de façon équilibrée et impartiale. Ce sont des spécialistes indépendants et reconnus dans le domaine du programme évalué. Il est demandé aux experts d'examiner toutes les données et toutes les analyses effectuées dans le cadre de l'évaluation, puis de dégager un consensus sur les conclusions que l'évaluation doit en tirer, et en particulier sur les réponses à apporter aux questions évaluatives. Le panel n'explicite pas complètement ses références de jugement ni ses arbitrages entre critères, mais la crédibilité de l'évaluation est garantie par le fait que les conclusions résultent d'un consensus entre des personnes qui ont une forte notoriété et sont représentatives des différentes « écoles d'expertise ».

● Rédaction des conclusions et du rapport provisoire

L'indépendance de l'équipe d'évaluation lui confère toute latitude pour formuler les conclusions qu'elle pense pouvoir tirer des données collectées et analysées. Cette latitude est cependant encadrée par des contraintes de professionnalisme dont le respect s'impose.

En tout premier lieu, les conclusions sont construites et présentées sous forme de réponse aux questions évaluatives. Les réponses sont exprimées dans les termes même qui ont servi à poser la question dans le cahier des charges et ces réponses sont aussi conclusives que possible. Si une des questions posées ne reçoit pas de réponse, les raisons en sont explicitement énoncées. En plus de la réponse aux questions posées, l'équipe d'évaluation ajoute toutes les conclusions qu'elle considère comme fondées et pertinentes pour les responsables du programme. C'est particulièrement le cas si le travail a permis d'identifier des effets négatifs inattendus ou s'il a montré qu'un objectif a perdu sa pertinence.

Dans le cas de l'évaluation d'ensemble d'un programme, les conclusions sont formulées au niveau du programme, c'est à dire de façon transversale aux mesures ou aux actions contenues qui le composent. Ces conclusions transversales prendront par exemple les formes suivantes : les meilleures réussites en matière d'emploi ou d'environnement ont été obtenues dans telle partie du programme, telles mesures sont particulièrement efficaces,

telles mesures sont devenues moins pertinentes à la suite des changements du contexte, etc. Il n'est pas acceptable que l'évaluation d'ensemble du programme aboutisse à une juxtaposition de conclusions spécifiques à chaque mesure. Les Volumes 4 et 5 de cette Collection font des propositions techniques pour aider les équipes d'évaluation à formuler des conclusions transversales.

Une des règles du professionnalisme consiste à relier chaque conclusion, de la façon la plus explicite possible, aux données et aux analyses qui permettent de l'étayer : sources statistiques ou bibliographiques, faits et opinions collectés lors des enquêtes (en respectant l'anonymat lorsque cela est nécessaire), résultats d'analyses, références précises des comparaisons, etc. Les limites méthodologiques sont également discutées pour montrer si le lien entre les données et les conclusions est robuste et dans quelles limites les conclusions sont généralisables.

Il existe une différence de nature entre les conclusions et les recommandations. Les conclusions découlent des observations et des analyses effectuées par l'équipe d'évaluation. Les jugements de valeur qu'elles comportent reposent sur des références claires et explicites. Leur devenir appartient en dernier ressort aux responsables politiques du programme. L'indépendance de l'évaluation commande que les conclusions appartiennent en dernier ressort à l'équipe d'évaluation, même si elles peuvent utilement être discutées par le groupe de pilotage dans leur forme provisoire. Au contraire, les recommandations ont une nature plus subjective et plus politique. Il n'est pas choquant que les recommandations émanent conjointement de l'équipe d'évaluation et du groupe de pilotage. C'est une suggestion qui est développée dans la dernière partie de cet ouvrage (voir pages 198 à 200).

L'Encadré 34 présente le contenu type d'un rapport d'évaluation.

Encadré 34 - Contenu du rapport d'évaluation

Le contenu du rapport doit comprendre l'ensemble des éléments permettant l'appréhension claire des résultats de l'évaluation et la compréhension du raisonnement poursuivi. La richesse du rapport va obliger à présenter celui-ci en distinguant précisément les contenus selon leur nature. Six grandes parties vont se détacher :

- Les éléments d'information sur le programme évalué : l'historique du programme et des mesures qui le composent, l'analyse des documents, les objectifs affichés, etc.
- L'exposé méthodologique qui rappelle les différentes étapes du déroulement de l'évaluation, les choix arrêtés par le groupe de pilotage, les comptes-rendus des études réalisées et de leurs apports respectifs au questionnement évaluatif, les limites de la méthode employée.
- La présentation des informations qui constituent la base de l'évaluation : données secondaires ou primaires collectées.
- L'exposé des arguments, c'est à dire les analyses et interprétations qui ont permis de formuler des conclusions à partir des informations collectées.
- L'exposé clair et précis des conclusions qui doivent être des réponses aux questions posées dans le cahier des charges.
- Et, le cas échéant, si le cahier des charges le spécifie, les recommandations qui reflètent les opinions de l'équipe d'évaluation. Ces recommandations devront être clairement hiérarchisées en fonction de leur importance (de leur utilité).

De plus, tout rapport d'évaluation devra contenir un « *executive summary* », c'est à dire un résumé de deux à cinq pages maximum des six parties.

● Validation de l'évaluation et rapport final

Le commanditaire, ou plus souvent le groupe de pilotage, se prononce sur la validation du travail de l'équipe d'évaluation au stade du rapport final provisoire. Dans ce cas, l'appréciation donne lieu à l'une des trois décisions suivantes :

- L'équipe d'évaluation n'a pas rempli son contrat et doit effectuer un travail supplémentaire qui ne fera pas l'objet de contrat supplémentaire.
- L'équipe d'évaluation doit conduire de nouvelles études de terrain afin d'étayer empiriquement certaines de ses conclusions ou recommandations qui sont apparues intéressantes mais sur des sujets mal connus. Ce travail nécessitera l'établissement d'un contrat et d'un budget supplémentaires.
- Le rapport est accepté sur le plan technique et qualitatif.

Dans les deux premiers cas, l'avis du groupe de pilotage doit être confirmé par le commanditaire de l'évaluation. Ce n'est qu'après la validation des conclusions par les autorités compétentes que le rapport peut être définitivement adopté et les travaux clôturés. Cette étape est délicate car il s'agit de distinguer clairement la qualité technique du travail de l'équipe d'évaluation, d'une part, et les conclusions elles-mêmes, d'autre part. S'il est normal que le commanditaire vérifie la qualité technique de l'évaluation, il doit au contraire s'interdire d'interférer avec les conclusions en vertu du principe de l'indépendance de l'évaluation. Pour faciliter cette étape de validation, il est bon d'annoncer à l'avance les critères sur lesquels la qualité technique du travail sera jugée, par exemple en se référant à la grille proposée page 192.

Une fois le rapport validé, l'équipe d'évaluation et le groupe de pilotage peuvent classer les principales conclusions dans un tableau avec les annotations suivantes : « conclusion valide, crédible et solidement étayée », « conclusion peu étayée et nécessitant éventuellement des analyses plus approfondies », « conclusion conduisant à des recommandations intéressantes », « conclusion qui apporte une connaissance nouvelle », etc.

Pour conclure ces deux premiers chapitres, soulignons que la réalisation d'une évaluation d'ensemble d'un programme est indispensable dans la mesure où le principe de programmation est essentiel dans les règlements de Fonds structurels. Une telle démarche, à l'échelle du programme dans sa globalité, est aujourd'hui facilitée par l'expérience acquise aux cours des dernières années.

Une fois l'évaluation réalisée, la suite du processus consiste à diffuser le rapport et à organiser l'utilisation des conclusions qu'il contient ; étapes qui seront abordées plus loin (voir Partie IV). Avant de poursuivre, il faut cependant évoquer les approfondissements qui peuvent être envisagés en matière d'analyse d'impact. C'est l'objet des pages qui suivent.



Analyser les impacts macroéconomiques

En complément de l'évaluation d'ensemble d'un programme, il peut être nécessaire d'analyser les impacts macroéconomiques.

Les impacts des interventions des Fonds structurels au niveau d'une région ou d'un pays dépendent en premier lieu des impacts microéconomiques obtenus sur le terrain, mais ils dépendent aussi de mécanismes macroéconomiques qui n'apparaissent qu'à plus grande échelle.

Ces mécanismes sont difficiles à évaluer de façon crédible par des techniques simples. Il est intéressant et possible de progresser dans leur compréhension et d'en donner une représentation schématique. Leur estimation quantitative relève de techniques de modélisation qui ont fait l'objet, au cours des dernières années, de développements spécifiques pour l'évaluation de la politique européenne de cohésion.

Les impacts macroéconomiques à l'échelle d'une région ou d'un pays peuvent-ils être considérés comme la somme de tous les impacts microéconomiques ? Le débat n'est pas clos mais conduit cependant à reconnaître l'importance des mécanismes macroéconomiques qui conditionnent l'atteinte des objectifs globaux de développement. Cette section leur est entièrement consacrée.

Les quelques exemples de l'encadré ci-dessous permettent de distinguer ces deux types de mécanismes.

Encadré 35 - Exemples de mécanismes de niveau micro et macroéconomique

- Grâce à l'aide reçue, une entreprise se développe (niveau micro), elle achète plus à ses fournisseurs régionaux qui achètent eux-mêmes plus à leurs fournisseurs régionaux, etc. (effet fournisseur, niveau macro).
- Grâce à l'aide reçue, une entreprise se développe (niveau micro), elle distribue des salaires supplémentaires qui sont en partie dépensés dans la région, ce qui donne lieu à de nouveaux développements, puis à de nouvelles distributions de revenu, etc. (effet multiplicateur, niveau macro).
- Un investissement public remplit le plan de charge d'une entreprise régionale ce qui réduit la concurrence lors de la réalisation des investissements suivants, (niveau micro), ce qui fait augmenter le niveau général des prix, etc. (niveau macro).
- Pour financer leur contribution aux projets aidés (niveau micro), les autorités régionales augmentent les impôts, ce qui diminue les revenus disponibles, etc. (niveau macro).

Du point de vue de l'évaluation, les impacts microéconomiques peuvent être évalués en interrogeant un nombre limité de personnes, c'est à dire les destinataires directs (chez qui on observera des effets directs) et des personnes qui ont été en contact avec les destinataires directs (chez qui on observera des effets indirects). Il est indéniable que des effets importants se produisent au-delà de ces cercles limités de destinataires, comme le montre l'encadré précédent. Ces effets plus lointains impliquent des mécanismes macroéconomiques qui affectent potentiellement la globalité des marchés ou des acteurs économiques du territoire concerné.

■ Éviter les démarches trop simplifiées

Les évolutions macroéconomiques d'une région s'observent relativement bien à l'aide des indicateurs statistiques tels que le PIB par habitant ou le taux de chômage pour citer ceux qui sont le plus couramment employés. Cette disponibilité facile conduit naturellement les équipes d'évaluation à faire des constats d'évolution avant-après sur la situation économique régionale. Lorsqu'un programme représente quelques pour-cent ou fraction de pour-cent de l'investissement total régional, l'observation des évolutions globales n'a évidemment aucune signification en termes d'impact, pas plus que ne peuvent en avoir les comparaisons avec l'évolution économique globale des régions voisines, fussent-elles non éligibles aux aides publiques.

Ce n'est que dans le cas d'un transfert massif d'argent public vers une région en retard de développement que l'on peut fonder l'évaluation sur l'observation des grandes tendances économiques. C'est ce qui a été fait dès les années soixante pour l'évaluation des politiques régionales à l'aide d'une technique appelée « analyse structurelle-résiduelle » (*shift-share analysis*). Cette technique, ainsi que toutes celles qui sont présentées dans ce chapitre, fait l'objet d'une fiche descriptive dans le Volume 3 de cette Collection. Il s'agit d'une démarche d'évaluation « *top-down* » qui procède par comparaison entre les tendances d'évolution de l'emploi observables au niveau d'un pays et au niveau de la région aidée. L'analyse structurelle-résiduelle décompose l'emploi régional par grands secteurs et applique à ces secteurs les taux d'évolution observés au niveau national pour ces mêmes secteurs. On reconstitue de cette façon une situation contrefactuelle montrant comment la région aurait évolué si elle avait évolué comme la moyenne du pays. La comparaison avec l'évolution réelle fait apparaître une différence (le résidu) qui est entièrement imputé aux effets des aides extérieures. Le caractère évidemment très simplificateur de cette analyse ne se justifie que si les aides extérieures sont réellement massives.

Un certain nombre d'effets macroéconomiques peuvent s'estimer à l'aide de coefficients d'extrapolation tirés d'études ou de recherches antérieures, selon la même logique que celle qui a été décrite plus haut (voir page 135) pour l'estimation des impacts microéconomiques. Les coefficients d'effets macroéconomiques s'appliquent aux effets microéconomiques qui sont supposés avoir été estimés au préalable. Les formules sont du type :

$$\text{Effet macro} = \text{Effet micro} \times \text{Coefficient.}$$

Le coefficient d'effet fournisseur varie selon la taille et le degré d'ouverture de l'économie régionale. Le coefficient d'effet multiplicateur varie dans une fourchette comparable.

L'équipe d'évaluation doit s'entourer de sérieuses précautions, et si nécessaire, de conseils d'experts, avant d'employer des coefficients d'effet macro. Une des raisons qui pousse à cette prudence est que la valeur du coefficient varie beaucoup d'une région à l'autre. L'autre raison est que les mécanismes macroéconomiques sont nombreux et que l'on oublie souvent ceux qui sont défavorables. Ainsi par exemple, l'effet d'éviction ou l'effet de hausse des prix peuvent conduire à appliquer des coefficients négatifs importants.

Les limites inhérentes aux démarches les plus simples conduisent à proposer deux directions complémentaires pour l'évaluation des impacts macroéconomiques. La première consiste à développer un modèle logique des relations de cause à effet entre le programme et ses objectifs macroéconomiques. Cette démarche, de nature qualitative, insiste sur les effets qui sont les plus importants en termes de développement, à savoir les effets structurants. La seconde démarche, de nature quantitative, consiste à utiliser la modélisation économétrique pour simuler l'évolution de l'économie régionale avec et sans programme, afin d'en déduire une estimation d'impact. Ces deux démarches seront traitées dans les sections suivantes.

■ Articuler les effets aux niveaux micro et macro

Un programme de développement vise à produire des changements durables et structurels à l'échelle d'une région ou d'un pays. Les effets recherchés ont deux dimensions qui se définissent comme suit : ce sont des effets qui doivent perdurer même si l'aide publique vient à cesser (notion d'effet structurant) et ce sont des effets qui interviennent aux niveaux macroéconomique et macro-social.

Par opposition, les réalisations du programme sont toujours microéconomiques ou micro-sociales et datées dans le temps. Elles ne sont

pas appelées à perdurer indéfiniment, ce qui serait implicitement équivalent à un échec du programme. L'évaluation doit donc chercher à identifier et à articuler les chaînes de causes et d'effets qui permettent de faire la liaison entre des réalisations qui interviennent à une date donnée à des effets permanents, d'une part, entre les niveaux micro et macro, d'autre part.

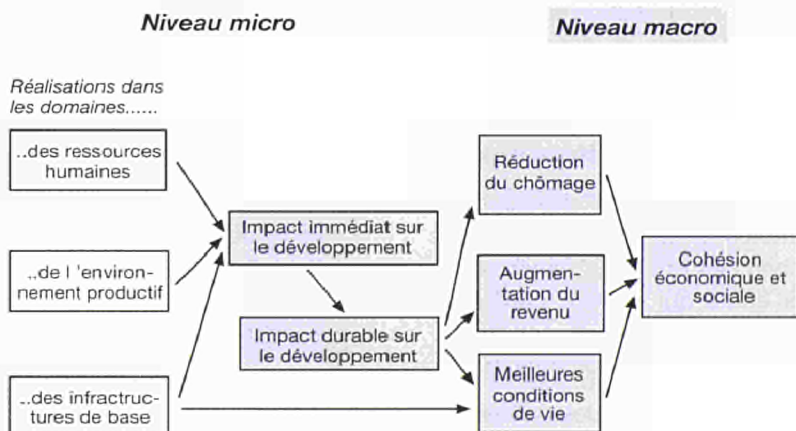
Cette démarche s'apparente à celle du diagramme logique des impacts qui est applicable aussi bien pour l'évaluation d'ensemble du programme (voir page 99), que pour les analyses approfondies (voir page 160).

● Impact global sur la cohésion économique et sociale

L'encadré ci-dessous donne une vue d'ensemble des mécanismes qui relient les réalisations des programmes des Fonds structurels à l'objectif global de la politique dont ils font partie, à savoir la cohésion économique et sociale. Ce schéma propose une représentation logique des effets des interventions sur le développement, à travers les mécanismes d'effets structurants. Les sections qui suivent vont permettre de parcourir ce schéma en le détaillant étape par étape.

Ces schémas ne peuvent être utilisés dans le cadre d'une étude particulière sans une nécessaire adaptation, il convient en effet que l'équipe d'évaluation et le groupe de pilotage les modifient en fonction de la stratégie du programme et du contexte régional.

Encadré 36 – L'articulation micro-macro : une vue d'ensemble



L'objectif de cohésion économique et sociale, tel qu'il est envisagé dans le cadre des politiques européennes et soutenu par l'ensemble des différents Fonds structurels, trouve ses bases dans l'article 130a du Traité sur l'Union européenne où il est défini en termes de « développement harmonieux » avec une dimension géographique spécifique : « réduire l'écart entre les niveaux de développement des diverses régions et le retard des régions les moins favorisées, y compris les zones rurales ». L'Agenda 2000 reprend cet objectif à part entière en lui adjoignant une dimension supplémentaire liée à l'élargissement de l'Union aux pays d'Europe de l'Est et aux objectifs de lutte contre le chômage, de promotion de l'égalité des chances entre les femmes et les hommes et de stimulation d'un développement compétitif et d'une croissance soutenable du point de vue de l'environnement.

● Impact des investissements dans les infrastructures de base

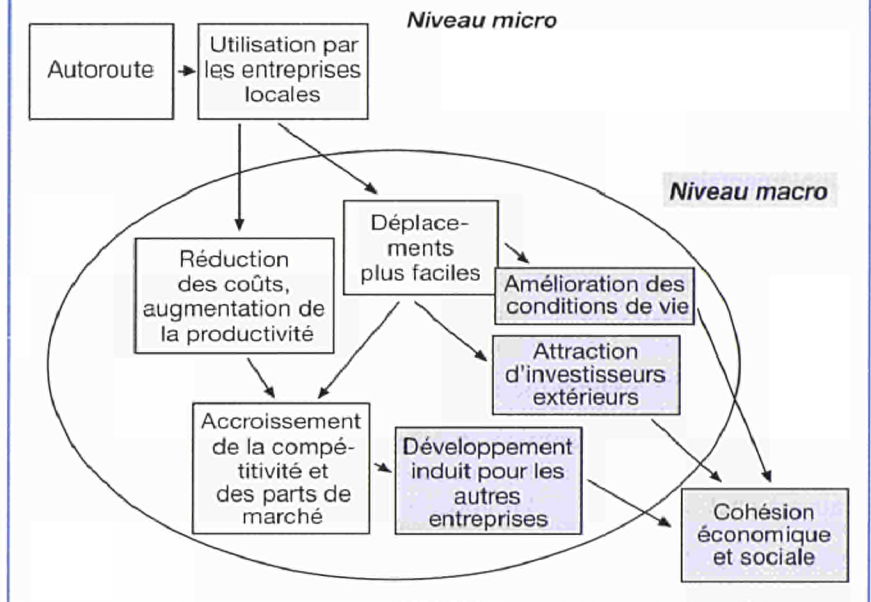
L'impact des infrastructures sur la cohésion est surtout indirect et profite aux entreprises et aux personnes de façon collective. Les relations de cause à effet par lesquelles les impacts se produisent sont multiples.

Exemples de projets d'infrastructure qui conduisent à des impacts durables au niveau macroéconomique :

- (1) Une nouvelle installation portuaire, une route, une autoroute ou un réseau de distribution d'électricité abaissent les coûts de production des entreprises régionales et les rendent plus compétitives.
- (2) Un système d'égouts ou un système de distribution de l'eau permettent de lever des contraintes de développement pour les entreprises locales, un aéroport rend possible le développement touristique.
- (3) Une amélioration des réseaux de télécommunications permet l'établissement de nouvelles relations entre les entreprises locales, ces relations améliorent le fonctionnement des réseaux d'innovation régionaux, ce qui améliore la compétitivité des PME régionales.

Les programmes d'infrastructures peuvent bénéficier aux activités économiques existantes mais aussi augmenter l'attractivité relative d'un territoire pour les investisseurs extérieurs. Les infrastructures peuvent aussi entraîner des bénéfices directs en termes d'accès aux biens publics (exemple : un temps de trajet plus rapide vers les hôpitaux).

Encadré 37 - Exemple d'effets structurants dans le domaine des infrastructures de base



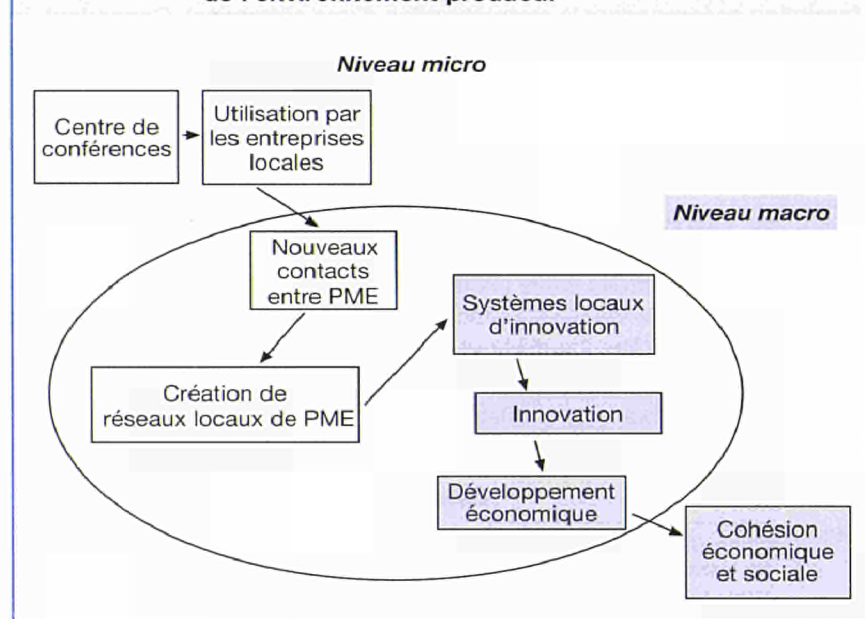
● Impact des interventions sur l'environnement productif

Les politiques régionales ont longtemps été fondées sur des aides directes aux investissements dits « de capacité ». Cette logique supposait que les investissements aidés permettent d'accroître la capacité de production de biens pour lesquels il existe une demande non satisfaite.

A présent, lorsque des aides sont directement accordées à des entreprises, les hypothèses qui sont faites sont souvent d'une autre nature. Il s'agit d'encourager ou de susciter des dépenses susceptibles d'accroître directement la compétitivité des entreprises aidées. L'idée est que le soutien européen aide les entreprises à bâtir des avantages concurrentiels. Ces avantages sont appelés « visibles » quand ils aboutissent à une meilleure « compétitivité-prix », grâce à des gains de productivité ou à des réductions de coûts de production. Les avantages « invisibles » ou immatériels sont de plus en plus déterminants dans la compétition mondiale et sont donc de plus en plus encouragés par les aides publiques. Il s'agit par exemple d'assistance à l'innovation, au marketing, à la qualité, etc.

Enfin, de plus en plus d'aides publiques cherchent à créer ou à renforcer des avantages compétitifs qui appartiennent collectivement à l'ensemble des entreprises d'une région ou d'une filière de production. L'environnement productif se compose de tous les facteurs que l'entreprise doit trouver en dehors de son organisation interne afin d'avoir de bonnes performances. Dans l'exemple de l'encadré ci-dessous, le programme subventionne un centre de conférence destiné à renforcer la compétitivité des entreprises régionales.

Encadré 38 - Exemple d'effets structurants dans le domaine de l'environnement productif



● Impact des interventions dans le domaine des ressources humaines

L'aide en faveur des ressources humaines a pour premier effet de renforcer et d'accroître les qualifications et les capacités des destinataires. Le concept de capital humain rend compte de l'accumulation des actions d'éducation et de formation « investies » dans les individus habitant la région. Ce concept constitue un pont entre la nature sociale des résultats d'une formation et ses effets macroéconomiques. Le capital humain regroupe à la fois les effets mesurables sur la productivité du travail et des

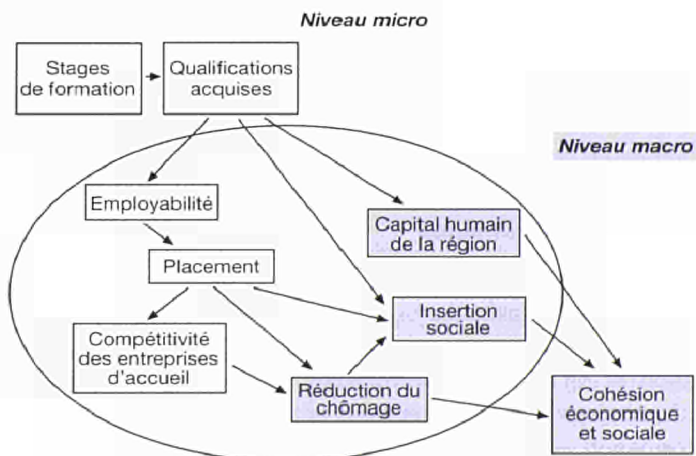
effets moins palpables sur l'adaptabilité, la pertinence des spécialisations ou la qualité des réseaux relationnels.

Les interventions structurelles en faveur des ressources humaines ont un impact direct sur les groupes concernés en augmentant leur employabilité, c'est à dire leur capacité à trouver un emploi ou à le conserver dans le contexte d'une région donnée et dans le cadre d'une conjoncture économique donnée.

L'aide en faveur des ressources humaines peut avoir un impact sur les entreprises (exemple : un programme de formation spécifique accompagne l'installation d'une nouvelle usine, un programme de formation accompagner la restructuration d'une entreprise). Cependant, les entreprises tirent essentiellement des avantages indirects sous forme d'une amélioration de l'adaptabilité et de la productivité de la main-d'oeuvre sur le marché du travail qui les concerne (exemple : une formation technique pointue peut renforcer les avantages d'une région déjà spécialisée dans un secteur dynamique et favoriser l'innovation).

Enfin, les interventions dans le domaine des ressources humaines ont souvent des impacts non-économiques essentiels tels que la reconstitution de liens sociaux détruits par le chômage et le retour à une intégration sociale et à un équilibre de vie. Ces impacts sont souvent liés au retour à l'emploi mais ils méritent d'être identifiés et si possible, évalués séparément.

Encadré 39 - Exemple d'effets structurants dans le domaine des ressources humaines



● Impacts macroéconomiques et stratégies de développement

L'objectif de cohésion suppose une réduction des disparités entre régions ou entre catégories sociales. Pour progresser vers cet objectif il faut admettre que les régions ou les catégories aidées doivent connaître un développement dont le rythme est supérieur au rythme moyen du développement économique européen. Pour cela, les programmes doivent non seulement stimuler le développement socio-économique mais aussi renforcer les moteurs de ce développement. Les différents types de stratégies qui sont adoptées à cet effet conduisent à attendre des impacts de nature très différente.

Le développement socio-économique peut être recherché par des moteurs externes telles que l'investissement, l'immigration de professionnels qualifiés ou le transfert de technologie. De nombreux programmes tentent d'attirer ces facteurs de développement. L'attractivité est le maître mot de ce type de développement appelé développement exogène. Une meilleure attractivité est un but recherché par toutes sortes de programmes dont les résultats s'expriment par exemple sous forme de baisse de prix des terrains, d'adaptation de l'infrastructure, de qualification de la main-d'œuvre, de qualité de vie urbaine ou de paysages protégés.

La délocalisation d'une entreprise vers un territoire aidé n'est pas en soi une preuve de succès en termes de développement car elle ne garantit pas un effet durable. Certains investissements peuvent en effet être attirés parce qu'ils sont mobiles. Cette mobilité est susceptible de perdurer au point que ces entreprises quitteront le territoire au bout de quelques années en direction de zones offrant des coûts de production plus faibles. En conséquence, la chaîne de causes et d'effets qui concerne le développement exogène doit être complétée par un dernier maillon : on doit supposer que les entreprises récemment installées créeront des liens suffisants avec la main d'œuvre et l'environnement productif du territoire de façon à rester durablement attachées à ce territoire.

Le développement socio-économique peut également être entretenu par des causes internes. Dans l'idéal, le développement endogène naît spontanément d'une base économique solide, c'est-à-dire d'un ensemble d'entreprises et de travailleurs fortement enracinés dans leur territoire, qui réussissent à s'imposer sur leurs marchés respectifs, et dont les marchés sont porteurs.

La force de la base économique d'un territoire peut s'appuyer sur sa spécialisation dans un secteur en forte croissance. Un programme qui vise à renforcer cette base pourra stimuler les secteurs d'activités stratégiques et dynamiques par des aides indirectes : formation, services collectifs, mise en réseau, accès à l'information, innovation, recherche, etc. L'aide au développement endogène suppose qu'un facteur de croissance préexiste sous la forme d'une spécialisation dans certaines activités dynamiques ou sous forme d'un avantage compétitif régional. Le programme soutient les activités concernées, accroît leur compétitivité et leur spécialisation et le dynamisme de ces activités s'en trouvent renforcés.

Le développement endogène d'un territoire dépend beaucoup de l'existence de réseaux locaux non-marchands. La recherche d'effets de réseau est à la base d'aides aux « technopôles » et aux « systèmes d'innovation ». Ils aboutissent à créer un environnement économique, social et politique capable de construire et de reproduire ses propres facteurs de réussite. Dans ces systèmes, les avantages concurrentiels essentiels ne sont pas détenus par des firmes individuelles. Ils ne s'appuient pas non plus sur les facteurs de production comme la main-d'œuvre ou le capital. Les facteurs clefs de la réussite du territoire sont détenus de manière collective et informelle par un réseau de firmes et d'acteurs locaux. Ce réseau constitue la principale source d'avantages compétitifs (technologie, innovations, savoir-faire, etc.). Les firmes profitent du réseau de manière formelle et informelle : associations professionnelles, aide des collectivités locales et régionales, réseaux informels, mobilité inter-firmes du personnel, canaux d'information parallèles, etc. Ces moyens d'interaction sont généralement indépendants des transactions de marché, ce qui signifie qu'ils correspondent à la définition du mot externalité (effet économique qui ne dépend pas d'une relation client-fournisseur).

Un programme qui vise à encourager le développement endogène ne doit pas être évalué en fonction d'impacts purement économiques. La création d'un environnement économique, social et politique susceptible de construire et de reproduire ses propres avantages concurrentiels implique toutes sortes de dimensions. Le terme « potentiel local » représente cet ensemble de facteurs qui gouvernent le renouvellement du processus de croissance : capacité d'innovation, dynamisme institutionnel, acceptabilité du développement, etc.

■ Réaliser une estimation quantitative

On a examiné plus haut (voir p 146) des démarches simplifiées d'estimation quantitative telles que l'analyse structurelle-résiduelle (*shift-share*) et l'emploi de coefficients. Ces démarches sont intrinsèquement limitées dans leur capacité d'analyse. Une démarche théoriquement idéale consisterait à réaliser une modélisation quantitative de l'ensemble des mécanismes macroéconomiques telle qu'elle a été ébauchée de manière qualitative dans les pages précédentes. Cette démarche n'est probablement pas complètement réalisable mais des efforts considérables ont été faits dans cette direction.

Les techniques de modélisation permettent de représenter le fonctionnement simplifié de l'ensemble d'une économie et des différents marchés qui la composent (ex : marché des produits, marché du travail, marché de la monnaie, etc.). Un modèle schématise le fonctionnement de l'économie à l'équilibre, mais il est aussi capable de simuler l'évolution de cette économie à la suite d'un choc exogène (ex : variation du prix du pétrole, variation des taux de change, prélèvement fiscal), le rajustement des variables modélisées et l'atteinte d'un nouvel équilibre des marchés. Il est donc possible d'utiliser un modèle à des fins d'évaluation en considérant que le programme provoque un choc exogène (un impact primaire microéconomique) et en simulant les répercussions de ce choc dans l'ensemble de l'économie. Le nouvel équilibre atteint prendra en compte l'ensemble des mécanismes macroéconomiques, et notamment les effets structurants, pour peu que ces mécanismes aient été modélisés.

En pratique, on simule un scénario avec intervention publique et un scénario sans intervention publique, l'impact macroéconomique est estimé par différence entre les deux scénarios. Les modèles sont entièrement quantifiés et fournissent donc une estimation de l'ampleur des impacts sur toutes les variables macroéconomiques qui ont été modélisées.

◀▶ L'estimation des effets peut intervenir ex ante. Dans ce cas, on compare deux scénarios prévisionnels (avec et sans programme). Bien que cela soit plus innovant, on peut aussi employer un modèle dans une situation d'évaluation ex post. Dans ce cas, la comparaison porte, d'une part, sur l'évolution économique passée telle que le modèle l'a recalculée de façon approchée et, d'autre part, sur un scénario sans intervention publique, reconstitué ex post par le modèle.

Les modèles macroéconomiques disponibles se limitent généralement au niveau national. Les modèles construits à l'échelle régionale ou locale sont

rare, notamment parce que les flux économiques sont très vite dilués hors du territoire concerné. La modélisation garde cependant du sens pour des petites économies à condition qu'elles soient relativement fermées. C'est pourquoi les économies insulaires sont souvent équipées de modèles macroéconomiques. Par ailleurs, l'emploi d'un modèle est d'autant plus intéressant que l'intervention du programme atteint une masse suffisante par rapport à l'économie modélisée. On considère que pour atteindre cette masse critique, l'importance économique du programme doit être au moins de l'ordre de 1% du PIB de l'économie du territoire.

Dans le cadre des interventions structurelles européennes, l'expérience d'évaluation macroéconomique la plus avancée a été réalisée avec le modèle HERMIN. Ce travail est présenté de façon plus détaillée dans le Volume 3 de la Collection, en parallèle avec deux autres expériences. Le modèle HERMIN a été utilisé dans le cadre d'évaluations en Espagne, en Irlande et au Portugal. Ces évaluations ont introduit une innovation intéressante en simulant explicitement les effets d'offre, notion qui n'est pas éloignée de celle des effets structurants évoquée plus haut (voir page 147). Dans le cas des modèles HERMIN, on a simulé l'effet des programmes sur la productivité du travail et/ou des autres facteurs de production, ainsi que sur l'attraction des investissements étrangers et l'accroissement de la compétitivité des exportations.

Cette expérience constitue un exemple intéressant des travaux de recherche et de développement qui portent actuellement sur la modélisation des impacts macroéconomiques en termes de développement. Ces travaux cherchent à dépasser les limites des premières expériences de modélisation qui ne simulaient que les effets de demande. Les effets de demande résultent du fait que l'argent public qui est dépensé dans le cadre du programme, continue de circuler dans l'économie régionale sous forme de consommation des entreprises et des ménages. Ils ne prennent pas en compte les gains de productivité et de compétitivité. Il est clair qu'une approche réduite aux effets de demande serait trop limitée car elle supposerait implicitement que le programme n'a plus d'effets lorsque la dépense publique s'arrête. Ceci est évidemment contraire aux hypothèses de base de l'aide au développement socio-économique. C'est pourquoi les travaux de modélisation des effets d'offre seront particulièrement intéressants à suivre au cours des prochaines années.

Réaliser des analyses plus approfondies

En complément de l'évaluation d'ensemble du programme, qui doit traiter un grand nombre d'impacts avec ses propres méthodes d'analyses, il est possible et utile de réaliser des analyses spécifiques à propos de quelques impacts particuliers. Ces analyses pourront utiliser tout l'arsenal des techniques d'évaluation, collecter des données primaires et aboutir à des estimations quantitatives crédibles sur l'atteinte des objectifs du programme.

L'approfondissement de l'analyse implique la réalisation d'une enquête pour observer les changements intervenus chez les destinataires, enquête suivie d'une analyse de causalité qui vise à déterminer quelle part de ces changements peut être attribuée au programme.

▶▶ A la différence de l'évaluation d'ensemble d'un programme qui vise à être exhaustive, l'approfondissement d'une question concerne une catégorie de réalisations, un public et un mécanisme d'apparition des impacts. L'évaluation d'ensemble d'un programme peut s'accompagner de l'approfondissement d'une ou plusieurs questions. On peut aussi décider de lancer une évaluation uniquement consacrée à l'approfondissement d'une question. La notion d'évaluation thématique évoque le traitement d'une question mais elle n'est pas équivalente. Une évaluation thématique examine en effet comment un même impact (ex : compétitivité des PME) est obtenu par plusieurs sortes d'interventions, ou comment les interventions financées dans un domaine précis (ex : la R&D) produisent plusieurs sortes d'impacts. Les évaluations thématiques sont décrites dans le Volume 5 de la Collection. Elles sont trop complexes pour entrer dans la logique simple de cette section qui se résume à l'évaluation d'une intervention précise dont on attend un impact principal (ex : impact de la R&D sur la compétitivité des PME).

La question traitée, ainsi simplifiée, permet d'avoir recours à toutes les techniques classiques de l'évaluation, notamment pour l'observation des comportements des destinataires et pour l'attribution au programme de sa juste part des changements observés. L'approfondissement portant essentiellement sur l'observation des effets sur le terrain, ces analyses concernent donc l'évaluation à mi-parcours et l'évaluation ex post.

■ Approfondir l'analyse de la logique interne du programme

L'évaluation d'ensemble d'un programme suppose d'expliciter sa logique interne, et en particulier d'examiner la bonne articulation de ses objectifs aux différents niveaux (voir p. 99). La logique d'un programme peut se schématiser sous la forme d'un arbre d'objectifs, son objectif global constituant le tronc et ses multiples réalisations formant les rameaux. Si l'on analyse une catégorie de réalisation et un impact, cela revient à isoler une branche de l'arbre. Il est alors possible d'examiner cette branche beaucoup plus en détail. C'est ce que vont montrer les sections suivantes.

● Cadre logique

La technique du cadre logique, originellement conçue pour l'évaluation ex ante des projets dans les pays en développement, est un outil qui permet d'entrer simplement et rapidement dans la description des liens logiques entre une réalisation et un impact attendu. La technique consiste à produire une présentation matricielle qui énonce les attentes en matière de réalisations (Cf. en bas à gauche de l'Encadré 40) ainsi que pour les principales étapes logiques séparant les réalisations de l'impact global (colonne de gauche). La matrice du cadre logique décrit donc une branche de l'arbre des objectifs en plaçant le « rameau » en bas (réalisations) et le tronc en haut (impact global).

A chaque étape logique, l'équipe d'évaluation mentionne les moyens prévus ou à prévoir pour vérifier l'atteinte des objectifs (indices de succès vérifiables). Elle définit en même temps ce qui doit être considéré comme un succès ou un échec (réalisations, résultats et impacts suffisants / insuffisants). En obligeant à remplir cette seconde colonne de la matrice avec des indicateurs ou des descriptions qualitatives précises, la technique apporte souvent une clarification très utile qui permet de revenir à une description plus précise de la première colonne. La troisième colonne, intitulée « conditions et risques » fait une liste des facteurs exogènes qui peuvent faciliter ou contrarier le passage d'une étape à l'étape suivante. Ce point sera détaillé plus loin (Cf. le « modèle explicatif »). On trouvera plus de détails sur le cadre logique dans le Volume 3 de cette Collection consacré aux techniques d'évaluation.

Encadré 40 - Cadre logique

	Enoncé des attentes	Indices de succès vérifiables	Conditions et risques
Impact global	Développement économique et social		
Impact spécifique	Economie pour les entreprises régionales		
Résultat	Gain de temps sur parcours-type		
Réalisation	Amélioration des routes		

Il est recommandé de mettre en œuvre la technique du cadre logique de façon participative en impliquant les gestionnaires et les opérateurs chargés des réalisations évaluées.

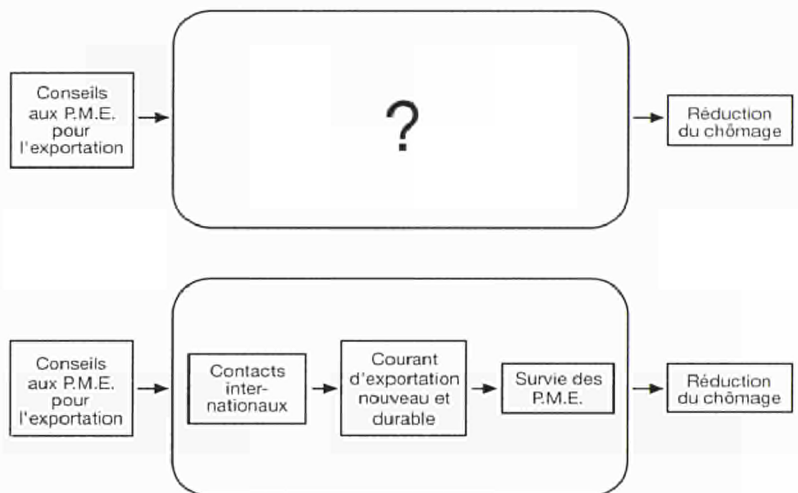
La colonne la plus à gauche de la matrice du cadre logique décrit la succession des effets du programme en quatre étapes (réalisations, résultats, impact spécifique, impact global). La colonne la plus à droite, en rappelant tous les facteurs extérieurs qui peuvent conditionner le passage d'une étape à l'autre, montre bien que l'obtention des impacts repose sur des hypothèses. C'est la raison pour laquelle les spécialistes de l'évaluation appellent cette colonne de droite « théorie d'action ». La théorie d'action est constituée de l'ensemble des hypothèses par lesquelles on explique qu'une réalisation donnée va produire un impact donné. L'évaluation a pour vocation première de confronter cette théorie à la réalité.

● Diagramme logique des impacts

Dans le cadre d'une analyse approfondie, les quatre lignes de la matrice du cadre logique ne suffisent généralement pas à détailler la théorie d'action de façon suffisamment explicite. En particulier, il subsiste souvent des zones d'ombre entre les résultats et les impacts. C'est ce que l'on a

coutume d'illustrer par l'image de la boîte noire des hypothèses d'impact. Cette image évoque la situation fréquente dans laquelle les concepteurs et les opérateurs savent bien expliquer ce qu'ils réalisent et l'impact global qui est espéré, mais ne savent pas expliquer de façon convaincante le processus qui est supposé conduire à cet impact.

Encadré 41 – Approfondir le diagramme logique des impacts



L'exemple de l'encadré ci-dessus propose de décomposer les hypothèses d'impact en quatre étapes. Entre chacune de ces étapes, il peut être nécessaire de rechercher des impacts intermédiaires et d'affiner le diagramme logique des impacts. En particulier, le diagramme mérite d'être détaillé chaque fois qu'une relation de cause à effet paraît obscure ou mal justifiée. Cela nécessite des échanges intenses avec les gestionnaires et opérateurs, au cours desquels l'équipe d'évaluation révèle leurs hypothèses implicites et produit une théorie complète en cherchant à éliminer progressivement les faiblesses logiques des raisonnements.

Lorsque le groupe de pilotage construit lui-même le diagramme logique des impacts, l'expérience montre qu'il considère cette étape comme une valeur ajoutée significative de l'évaluation.

L'exemple de l'encadré précédent montre que l'ouverture de la boîte noire des hypothèses implicites permet de diriger le travail d'évaluation vers

certaines hypothèses d'impact plutôt que d'autres moins stratégiques. Cet exemple montre que l'évaluation a peu de chances de conclure si l'observation se limite au nombre de jours de conseil fourni aux PME d'une part et aux statistiques du chômage d'autre part. Par contre, l'analyse d'impact sera plus intéressante si on demande aux responsables des PME aidées combien de contacts internationaux ont été générés ou facilités par le programme, comment ces contacts ont évolué, combien de ces contacts sont encore actifs, quel chiffre d'affaires ils ont généré, quel pourcentage ils représentent dans l'activité de l'entreprise, etc. L'analyse approfondie de la logique du programme permet de focaliser l'observation de terrain sur le bon niveau d'impact.

■ Réaliser les observations sur le terrain

La dimension empirique et pragmatique d'une évaluation repose sur sa capacité à observer les effets du programme sur le terrain. C'est donc à travers des enquêtes auprès des destinataires du programme que l'évaluation apporte toute sa valeur ajoutée. La réponse approfondie à une question évaluative comprend donc souvent une collecte de données primaires auprès des personnes ou organismes qui ont participé au programme, et également auprès de non-participants quand l'évaluation use de techniques comparatives évoquées par ailleurs (voir pages 116 et 166). Les sections qui suivent décrivent les principales techniques d'enquête utilisées en évaluation.

● Collecte de données quantitatives

L'enquête par questionnaire est le moyen le plus classique de collecter des informations quantifiables. Elle peut être réalisée par courrier, par téléphone ou par entretiens. La formulation du questionnaire est une étape décisive. Les questions sont en nombre limité, énoncées de manière claire et formulées de façon à intéresser la personne interrogée. Le questionnaire propose des questions fermées auxquels les personnes interrogées peuvent répondre en choisissant parmi des réponses proposées par le questionnaire ou l'enquêteur. On peut éventuellement utiliser quelques questions ouvertes, mais les réponses doivent alors être codées (classées par catégories) après collecte des données, ce qui est très lourd. Un pré-test du questionnaire auprès d'un nombre limité de personnes est indispensable afin de valider la forme et le contenu du questionnaire et de

confirmer sa capacité à collecter l'information. L'organisation d'une enquête peut nécessiter plusieurs mois et représente un coût variable en fonction de la taille de l'échantillon, de son accessibilité, du taux de réponse souhaité, etc.. Les différentes enquêtes entreprises dans une région doivent également être coordonnées avec soin pour ne pas provoquer de réactions de rejet des destinataires (les refus de participer à une enquête sont de plus en plus fréquents).

Les enquêtes par questionnaire sont largement utilisées à toutes sortes de fins mais leur domaine d'emploi privilégié est celui du recueil d'information quantitative auprès des destinataires. Grâce au caractère fermé des questions posées, les données fournies par un questionnaire peuvent en effet donner lieu à la construction d'indicateurs. Un fondement de l'enquête par questionnaire est d'utiliser un échantillon de telle sorte que des conclusions puissent être généralisées à l'ensemble de la population. Il suffit souvent d'un échantillon restreint (par exemple 200 à 300 stagiaires) pour obtenir la quantification correcte d'un indicateur.

La taille de l'échantillon doit cependant être plus importante si l'on désire exploiter pleinement le caractère quantitatif de l'information fournie par une enquête et en particulier lorsque l'on veut comparer les résultats de deux enquêtes, comme dans les trois exemples suivants : une enquête auprès des participants à une formation sachant qu'une certaine proportion d'entre eux occupe un emploi après 12 mois, et que l'on veut faire une comparaison avec un groupe de non-participants ; une enquête auprès des entreprises aidées sachant qu'un petit nombre d'entre elles développent des projets de haute technologie, et que l'on veut faire une comparaison avec les résultats d'une enquête nationale ; une enquête sur le revenu des agriculteurs sachant qu'une partie de celui-ci provient du tourisme, et on souhaite faire une comparaison avec les résultats d'une enquête réalisée il y a cinq ans sur un échantillon partiellement identique. Dans chacun de ces exemples, l'équipe d'évaluation veut comparer les résultats d'enquêtes portant sur des échantillons différents ou partiellement différents. Dans ce cas, les règles de validité statistique interdisent les comparaisons si les échantillons ne sont pas suffisamment importants. La taille minimum varie de 500 à 2000 individus en fonction de critères qui ne dépendent pas de la taille de la population enquêtée. L'appel à un professionnel permet de choisir la taille correcte d'un échantillon dans une situation concrète. A défaut, on retiendra que la taille des échantillons statistiquement comparable est le plus souvent comprise entre 900 et 1100 individus.

Une forme particulière d'enquête est celle qui permet de suivre les individus ou groupes d'individus au cours du temps et de mesurer ainsi les changements d'attitude ou de situation. On utilise généralement ces études par panel dans le cas d'interventions dans le domaine de l'emploi (on interroge les participants avant et après une formation par exemple). Ces enquêtes s'étendent sur plusieurs années et peuvent donc difficilement être lancées dans le cadre d'une évaluation dont la durée est limitée à quelques mois. Cependant, lorsque les enquêtes par panel existent, l'évaluation utilise leurs résultats en tant que données secondaires.

● Techniques d'enquête qualitative

Une analyse approfondie combine généralement la collecte d'informations quantitatives et qualitatives. Plusieurs techniques d'enquête qualitative sont présentées dans le Volume 3 de la Collection consacré aux outils d'évaluation. L'entretien individuel non directif et l'étude de cas (voir page 134) sont les outils les plus courants. L'équipe d'évaluation ne doit cependant pas enfermer ses choix techniques dans les limites de l'habitude et c'est pourquoi on évoquera ci-dessous deux techniques qui peuvent se révéler très concluantes.

Le premier exemple est celui de l'entretien de groupe (*focus group*). Sous réserve qu'il soit animé par une personne qualifiée, l'entretien de groupe permet de créer une dynamique d'échange entre les participants, de faire jouer leur créativité, et de recueillir finalement beaucoup plus d'information que ce qui résulterait d'entretiens individuels avec chacun des participants. L'entretien de groupe est particulièrement indiqué lorsque les destinataires ont une opinion très vague des effets que le programme a produit pour eux (cas d'une PME qui a participé à un réseau technologique par exemple). Leurs réponses à un questionnaire ou lors d'un entretien individuel risqueraient d'être des réponses de politesse. L'entretien de groupe permet dans ce cas d'obtenir des réponses plus fiables.

Le second exemple est plus atypique, c'est celui de l'observation ethnographique qui implique que l'enquêteur s'immerge parmi les participants au programme (en tant que participant ou assistant à un stage d'insertion par exemple). Cette technique a été appliquée avec succès pour des enquêtes auprès de populations difficiles, peu motivées, voire réticentes.

■ Analyser les impacts

Les analyses approfondies sont celles qui permettent d'atteindre les niveaux les plus élevés dans l'étude des impacts (voir Encadré 29 page 123). C'est pourquoi l'analyse d'impact est particulièrement développée dans les sections suivantes.

L'analyse des impacts combine deux types de démarches dans des proportions qui varient selon la situation : l'analyse inductive et l'analyse déductive. L'analyse inductive suggère des liens de causalité à partir de l'observation des faits. Elle est donc exploratoire dans un premier temps et vise à produire une théorie explicative des causes et des effets en partant d'observations, le plus souvent qualitatives, effectuées sur le terrain. L'analyse déductive considère au contraire cette théorie explicative comme un point de départ qu'il s'agit de tester par confrontation à la réalité. En cas de validation des hypothèses initiales, elle peut conduire à une estimation quantitative des impacts. Les deux sections qui suivent expliquent respectivement comment construire une théorie explicative, comment estimer les impacts, quand utiliser la théorie explicative et quelles sont les limites à l'analyse d'impact. Il s'agit d'une convention car une évaluation peut procéder de façon inverse si les circonstances le justifient.

● Théorie explicative des impacts

A plusieurs occasions, cet ouvrage a déjà fait mention des relations de cause à effet entre le programme et ses impacts et de la manière dont ces relations peuvent être schématisées sous forme d'un diagramme logique des impacts. La théorie explicative ne constitue qu'un pas supplémentaire dans la représentation du processus de production des impacts, mais c'est un pas difficile et décisif pour l'évaluation.

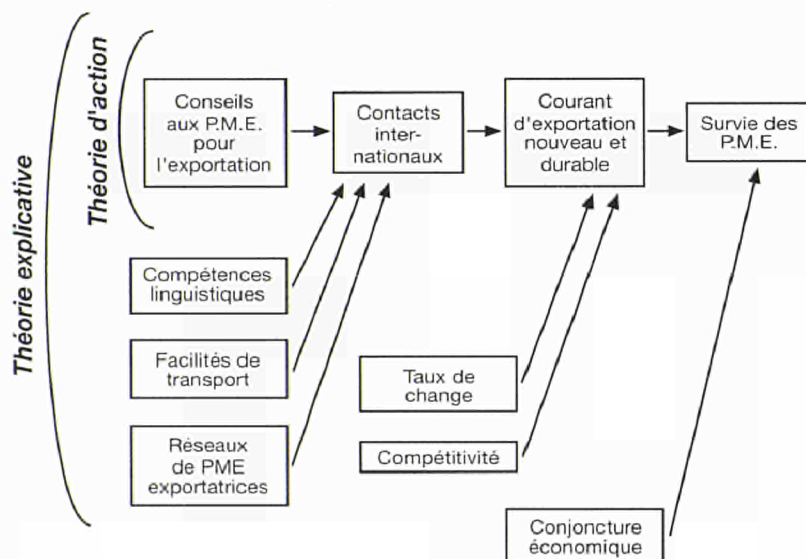
L'encadré 42 ci-après montre que l'on passe du diagramme logique à la théorie explicative en ajoutant tous les facteurs qui peuvent faciliter ou contrarier l'obtention des impacts à chaque niveau. On retrouve ici ce que la technique du cadre logique appelle « conditions et risques » et ce que la littérature de l'évaluation appelle « facteurs exogènes » ou « facteurs concurrents ». Dans toute situation d'évaluation, les facteurs exogènes sont en nombre potentiellement infini. Le rôle de l'analyse inductive est de déterminer ceux qui apparaissent les plus influents au regard de ce qui a été observé sur le terrain, par enquête auprès des destinataires.

Dans le cadre de son analyse inductive, l'équipe d'évaluation va également chercher à identifier tous les effets inattendus qui sont observables chez les

destinataires. Par exemple, plusieurs évaluations ont montré un effet négatif pour les habitants des zones aidées, découlant du fait que ces zones sont déclarées comme « défavorisées » et deviennent, en conséquence, moins attractives (effet de stigmatisation). Il existe heureusement aussi des effets inattendus positifs.

La technique de l'étude de cas présente un excellent rapport coût-efficacité pour déceler les facteurs exogènes et les effets inattendus. C'est probablement l'outil de base de l'analyse inductive. On peut également apprendre beaucoup à l'aide d'une série d'entretiens individuels avec des personnes ayant participé et n'ayant pas participé au programme. Il ne faut cependant pas considérer que l'analyse inductive est toujours qualitative. Des techniques comme l'analyse de variance ou l'analyse factorielle sont également utilisées de façon courante pour révéler les facteurs exogènes les plus influents. Sans aller vers des outils aussi sophistiqués, des tableaux, des cartes et des graphiques bien choisis peuvent souvent révéler des coïncidences et similitudes qui suggèrent de façon très parlante l'importance de certains facteurs exogènes.

Encadré 42 - Théorie d'action et théorie explicative



● Estimation des impacts

L'analyse d'impact consiste à attribuer une part des changements constatés au programme et une autre part aux facteurs exogènes, ou tout au moins aux principaux facteurs exogènes. Ces derniers entrent en concurrence avec le programme pour expliquer les différences « avant-après », d'où leur appellation de facteurs concurrents.

Exemple : Analyse d'impact à l'aide d'un groupe apparié

Une analyse comparative, par groupe apparié, a été menée au Royaume Uni, dans le cadre de l'évaluation d'un programme qui s'adressait aux PME en leur offrant des prestations de conseil en gestion, de formation et d'information. Les impacts de ce programme ne pouvaient pas être exprimés en emplois créés, puisqu'il visait à augmenter la productivité, ce qui conduisait à court terme à détruire des emplois. Il a donc fallu évaluer le programme selon d'autres critères comme la productivité, la rentabilité ou l'investissement induit.

Un échantillon d'environ 300 entreprises destinataires a été constitué. Un autre échantillon d'environ 200 entreprises non destinataires a également été constitué selon la technique des binômes appariés. Pour assortir les binômes, les critères suivants ont été utilisés : type d'activité de l'entreprise, taille de l'entreprise mesurée en effectif, ancienneté et situation géographique.

Un questionnaire a été adressé aux entreprises destinataires et à celles du groupe apparié. Une comparaison statistique entre les entreprises des deux groupes a permis d'obtenir des indications quantitatives sur les effets du programme en termes de chiffre d'affaires, de productivité et d'investissements induits. Afin d'éviter des déclarations "biaisées", les questionnaires comportaient des questions nombreuses et partiellement redondantes (questionnaire triangulé). L'enquête intégrait la recherche d'effets d'aubaine.

L'estimation d'impact nécessite d'estimer l'effet des facteurs exogènes, puis de soustraire cet effet des changements constatés puis enfin d'attribuer au programme la différence. La « situation contrefactuelle » est celle qui résulterait de l'influence des facteurs exogènes et d'eux seuls. Pour estimer la situation contrefactuelle, la technique la plus fréquente consiste à reconstituer un groupe de comparaison avec des personnes qui n'ont pas participé au programme, par exemple, des candidats dont le projet a été refusé, des candidats qui ont abandonné le programme en cours de route, des non-candidats ayant un profil similaire aux participants (groupes appariés comme dans l'exemple précédent). On peut aussi

reconstituer un groupe de comparaison fictif à partir d'individus ayant répondu à une enquête nationale (comparaison statistique par redressement). Une fois le groupe de comparaison construit, différents dispositifs de comparaison peuvent être mis en oeuvre : une comparaison « après seulement », une comparaison « avant/après » ou l'utilisation de séries temporelles regardant avant, pendant et après. Pour une description détaillée de la technique, se reporter au Volume 3 de la Collection.

L'équipe d'évaluation doit composer le groupe de comparaison jusqu'à ce qu'il ressemble suffisamment à celui des destinataires directs du programme. Les deux groupes doivent avoir les mêmes caractéristiques essentielles, ce qui revient à dire qu'ils doivent être influencés par les mêmes facteurs exogènes. En effet, des différences entre ces facteurs seraient susceptibles d'expliquer une part des différences constatées entre les deux groupes et cela fausserait l'estimation d'impact. La construction ou le redressement du groupe de comparaison suppose donc que l'on ait une bonne connaissance des facteurs exogènes, c'est à dire que l'on ait une théorie explicative des impacts.

● Utilisation de la théorie explicative

La théorie explicative est encore plus nécessaire quand l'estimation d'impact fait appel à une analyse de régression. L'emploi de cet outil commence en effet par l'écriture d'un petit modèle qui schématise les causes (le programme et les principaux facteurs exogènes), les effets (le changement constaté) et les relations de cause à effet. Ce modèle est ensuite ajusté de façon à représenter le plus fidèlement possible les données disponibles (données de gestion et/ou données d'enquête). Le résultat de l'analyse est une estimation quantitative de l'impact. La validité de l'estimation dépend fortement de la qualité du modèle de départ, et notamment du fait que ce modèle prend correctement en compte les facteurs exogènes.

L'analyse de régression est un exemple typique d'analyse déductive. La technique nécessite de formaliser dès le départ un ensemble d'hypothèses sur les causes et les effets (pour cette raison, on parle aussi d'analyse hypothético-déductive). Elle utilise pour cela une théorie explicative empruntée à la littérature scientifique, ou puisée dans le savoir d'un expert, ou encore construite par analyse inductive dans le cadre de l'évaluation. Dans un second temps, les hypothèses sont confrontées à la réalité et

l'équipe d'évaluation conclut sur l'existence de l'impact, sur son importance quantitative, ou sur sa non-existence.

● **Limites de l'analyse d'impact**

L'évaluation d'un impact est un exercice qui s'apparente à une prévision, même lorsque la démarche est rétrospective. Que serait-il arrivé si l'argent public n'avait pas été dépensé ? Il n'y a pas d'évaluation ex post sans réponse à cette question. Or la description de ce qui serait arrivé sans l'action publique, la construction de la situation contrefactuelle, n'est rien d'autre qu'une prévision effectuée à rebours. L'évaluation partage donc toutes les limites de la prévision. Et on sait qu'elles sont nombreuses.

Certains contextes se prêtent facilement à la prévision et à l'évaluation alors que d'autres sont tellement complexes qu'aucun des deux exercices n'aboutit à des résultats crédibles. Dans un contexte imprévisible, l'équipe d'évaluation ne peut pas établir la situation contrefactuelle de façon convaincante, et même les participants au programme sont incapables d'estimer l'impact qui s'est produit dans le cadre de leur propre projet.

Les situations favorables sont celles où la théorie explicative est relativement simple et linéaire. Par exemple, des investissements publics améliorent un aéroport, ce qui entraîne directement la multiplication des dessertes aériennes, laquelle conduit directement à une augmentation de la fréquentation touristique. Dans cet exemple, la cause est clairement connue et conduit mécaniquement à un effet relativement prévisible. Il est possible d'estimer les impacts dans une telle situation, quitte à y mettre le prix en termes d'enquêtes et d'analyses.

Les situations défavorables sont celles où la théorie explicative est complexe, particulièrement lorsqu'elle contient des relations de causalité circulaire. Ainsi en est-il de l'impact des programmes de formation dont les effets suivent des chemins embrouillés. D'un côté, ils produisent de nouvelles qualifications qui doivent s'ajuster aux besoins économiques afin d'améliorer l'employabilité des stagiaires et d'aboutir à de bons taux de placement. D'un autre côté, ils fournissent aux entreprises une main d'oeuvre plus productive, ce qui accroît leur productivité mais diminue donc l'emploi. Toutefois les gains de productivité ont une influence favorable sur la compétitivité, ce qui permettent de créer ou de sauvegarder indirectement des emplois. L'analyse d'impact dans ce contexte devient très difficile car l'explication causale peut donner lieu à plusieurs versions

différentes : un bon taux de placement peut être soit la cause, soit la conséquence des gains de compétitivité.

La difficulté est encore plus évidente si l'on admet que ce phénomène de circularité s'applique également à la politique elle-même. On peut en effet espérer que les chômeurs sont aiguillés vers des formations où les taux de placement sont élevés plutôt que dans des directions aléatoires. Est-ce que cela ne signifie pas que la politique elle-même est influencée par les impacts qu'elle est censée produire ? Avec ce type de causalité circulaire, le programme évalué peut être regardé alternativement comme une cause ou comme un effet. Dans cette situation, les concepts d'effets, d'impacts et de résultats perdent leur sens.

Certaines analyses d'impact sont donc impossibles à réaliser à cause de la nature même du problème posé, quel que soit le professionnalisme de l'équipe d'évaluation. Le cahier des charges doit donc être rédigé avec prudence car certaines questions relatives aux impacts du programme résistent à l'évaluation, même comportant des analyses approfondies.



IV

Utiliser l'évaluation



Optimiser l'utilisation de l'évaluation

Pour qu'une évaluation ait une réelle valeur ajoutée dans la sphère décisionnelle et institutionnelle, il faut que ses conclusions soient correctement diffusées et qu'elles rencontrent un écho favorable auprès des utilisateurs potentiels.

Le support constitué par le rapport d'évaluation n'est pas nécessairement le seul ni le plus efficace en particulier lorsque les utilisateurs sont des élus ou des citoyens qui recevront les résultats par l'intermédiaire des médias. Ces considérations plaident fortement pour que soit établi suffisamment tôt un plan de communication des résultats de l'évaluation.

Au-delà de l'étape de diffusion, arrive la dernière et la plus importante des étapes de l'évaluation : celle de son utilisation. Tout le processus d'évaluation mérite d'être optimisé en vue d'obtenir la meilleure utilisation possible. Il convient cependant de ne pas être exagérément optimiste sur certaines formes d'utilisation. Si un bon rapport d'évaluation apporte toujours des connaissances nouvelles (utilisation cognitive), il ne conduit pas obligatoirement ses lecteurs à réviser leurs jugements dans le sens des conclusions (dimension normative). La prise en compte des recommandations de l'évaluation (valeur ajoutée instrumentale) est la plus délicate à obtenir. En effet, les processus de décision publique font souvent intervenir plusieurs acteurs et de très nombreux facteurs.

■ Diffuser les conclusions de l'évaluation

Dans les sections qui suivent, on appellera utilisateur toute personne qui a l'occasion de lire ou d'entendre les conclusions et les recommandations de l'évaluation, et pour qui cette information provoque un changement de connaissance, d'opinion ou d'attitude. Le spectre des utilisateurs potentiels de l'évaluation s'étend depuis les co-décideurs du programme, jusqu'au lecteur du journal ou de la revue qui relate l'évaluation, en passant par les opérateurs et les gestionnaires.

La diffusion de l'évaluation est une phase essentielle qui suit la production du rapport d'évaluation, mais qui mérite d'être prévue dès la phase de rédaction du cahier des charges. Il est souhaitable que le cahier des charges annonce à qui les résultats de l'évaluation seront diffusés et sous

quelle forme ils le seront. Les décisions en la matière sont confirmées par le commanditaire après la remise du rapport, mais une bonne anticipation de la stratégie de diffusion permet d'optimiser l'utilisation de l'évaluation.

● Les canaux de communication

La stratégie de diffusion comporte un compromis entre l'objectif de maximiser la visibilité et l'objectif d'établir un climat d'interaction confiante avec tous les protagonistes de l'évaluation. En ce qui concerne les enquêtes conduites par l'équipe d'évaluation auprès des destinataires et des opérateurs du programme, il suffit, pour assurer confiance et visibilité, de traiter tous les résultats d'enquête de façon confidentielle en préservant l'anonymat des personnes interrogées durant tout le processus d'évaluation.

En ce qui concerne les interactions entre l'équipe d'évaluation, le groupe de pilotage et le commanditaire, cette question peut être plus difficile à gérer. Les interactions portent en effet sur des points sensibles telles que l'explicitation des critères et des normes, l'identification d'exemples de bonnes et de mauvaises pratiques, l'importance à donner à telle question ou à tel critère plutôt qu'à tel autre, etc. Les questions sensibles ne peuvent pas être traitées correctement avec une totale publicité. Au contraire, l'équipe d'évaluation doit respecter une discrétion suffisante dans son travail et dans son rapport pour permettre aux membres du groupe de pilotage de s'exprimer sincèrement en leur nom propre et pour leur permettre de dépasser, si nécessaire, le discours officiel de leur institution. Dans certains cas, l'équipe d'évaluation rédige son rapport de façon à préserver un haut degré de discrétion sur ses interactions avec les responsables du programme, et elle rédige une note supplémentaire et confidentielle pour attirer leur attention sur des points plus sensibles.

Trop souvent, le problème de l'utilisation de l'évaluation se réduit à la question : « publier ou ne pas publier le rapport final ? » alors qu'un ensemble plus large de questions devrait être pris en considération. Au-delà de la question de la publication, c'est la façon de diffuser qui doit faire l'objet d'une réflexion et d'un choix. Diverses formes de communication et de nombreux supports possibles sont présentés dans l'encadré ci-après.

Encadré 43 - Canaux de communication*Canaux de communication*

Audience		Entretien réunion	Rapport provisoire	Note confidentielle	Rapport public	Synthèse	Présentation	Brochure article	Conférence de presse
	Commanditaire			★	★	★	★		
	Groupe de pilotage Gestionnaires Opérateurs	★	★		★		★		
	Institutions européennes, nationales, régionales				★	★		★	
	Citoyens Journalistes					★		★	★

Les personnes impliquées dans le programme (groupe de pilotage de l'évaluation, gestionnaires et opérateurs) ont tendance à utiliser l'évaluation dès que les premiers résultats sont présentés par l'équipe d'évaluation au groupe de pilotage (ex : résultats d'enquêtes et d'analyses, rapports intermédiaires, rapport final provisoire). De leur point de vue, l'utilisation de l'évaluation commence avant la communication du rapport final.

Pour les autres utilisateurs il est indispensable de prévoir une stratégie de communication adaptée combinant un support et une échéance. L'encadré qui précède montre que différents canaux de communication sont possibles qu'il convient de choisir en fonction de la plus ou moins grande disponibilité des différentes catégories d'utilisateurs (« récepteurs » selon les spécialistes de la communication).

● Plan de communication

Le commanditaire et les plus hautes autorités administratives et politiques, n'ont pas le temps de lire un rapport d'évaluation en détail. Le mode de communication le plus efficace à ce niveau est une courte synthèse de quelques pages, renvoyant à certaines sections du rapport final pour plus de détails. La communication peut utilement être complétée par une présentation orale par le groupe de pilotage ou son Président au cours d'un séminaire ou d'une réunion après l'achèvement du travail.

Pour élargir la communication aux Institutions européennes, nationales et régionales, on peut également utiliser un court document de synthèse. Cependant, la communication passe mieux si cette synthèse est transformée en une brochure contenant des exemples concrets de réalisations du programme et montrant comment ces réalisations ont été jugées par l'évaluation. L'information des partenaires sociaux et des destinataires du programme sera réalisée de façon optimale en publiant des documents spéciaux (bulletins, fiches de synthèses, etc.). D'autres documents seront préparés pour l'information du « grand public » (brochures, articles, etc.) et des médias. On peut à cet égard prévoir une conférence de presse à laquelle est associé le groupe de pilotage comme cela est fait parfois pour obtenir un impact auprès des journalistes d'autant plus fort que les membres du groupe renforcent la légitimité de l'évaluation.

Dans la plupart des cas, les règles de transparence administrative imposent de rendre public le rapport final. Malgré tout, la diffusion publique de ce rapport n'est qu'un moyen parmi d'autres, et pas toujours le plus efficace, pour faire parvenir les conclusions de l'évaluation à ses utilisateurs. Un effort sincère de publication complète et de large diffusion aura cependant l'avantage d'exposer l'évaluation à la critique des experts, ce qui est une garantie de qualité. Si le commanditaire annonce que le rapport sera publié et largement diffusé, la qualité méthodologique du rapport a toute chance d'être meilleure car l'équipe d'évaluation sera stimulée par cette perspective et l'évaluation sera plus crédible car organisée de façon plus transparente.

En revanche, il y a un risque certain d'autocensure : les conclusions et recommandations n'en seront que plus prudentes. Pour éviter cet inconvénient qui peut se révéler être un véritable handicap, on peut demander à l'équipe d'évaluation de produire une note confidentielle en plus de son rapport.

Une bonne communication aux utilisateurs de l'évaluation suppose par ailleurs de « phaser » le processus de diffusion. Si l'évaluation comporte des enquêtes de terrain et une analyse approfondie de certaines questions, il convient d'organiser des présentations partielles des résultats obtenus sans attendre l'achèvement du rapport final.

Le plan de communication doit surtout être soigneusement synchronisé avec les grandes échéances du programme : dates des principales réunions de gestion, échéances décisionnelles du programme, sortie des rapports publics sur les politiques concernées par le programme au niveau régional, national et européen.

En réalité, ce n'est pas seulement le plan de communication, mais tout le processus d'évaluation, qui mérite d'être organisé en fonction du calendrier des utilisateurs. Cette préoccupation est présente dès les tous premiers moments de la commande (voir page 68) et elle est déterminante pour l'utilité de l'évaluation.

■ Considérer toutes les formes d'utilisation

L'évaluation est fréquemment jugée à l'aune de ses propres critères. On lui demande, bien évidemment, d'avoir une efficacité propre : sa valeur ajoutée est-elle suffisante par rapport aux ressources budgétaires et humaines qui lui sont consacrées ? Pour répondre à ce type de questions, de nombreux travaux de méta-évaluation ont été réalisés et publiés. La plupart conduisent à la conclusion qu'il faut considérer plusieurs types d'utilisation, l'utilisation directe des recommandations de l'évaluation par les décideurs n'étant qu'une forme d'utilisation parmi d'autres. Trois grandes formes d'utilisation se dégagent.

● Utilisation pour élever le niveau des connaissances

◀▶▶ Toute évaluation va en premier lieu permettre à ses utilisateurs de découvrir, d'être parfois surpris, d'apprendre et si possible de comprendre. La phase de conception de l'évaluation aura un grand rôle à jouer à ce niveau puisqu'elle va contribuer à clarifier des objectifs qui étaient peut-être mal exprimés ou seulement implicites dans les documents de programmation, à expliquer les mécanismes d'impacts, à identifier les changements de comportement, à vérifier la réalisation des effets attendus et à mettre en avant les effets inattendus.

L'exercice d'évaluation transforme la manière dont l'utilisateur se représente le programme et ses effets. Cette forme d'utilisation peut être d'ordre conceptuel lorsque les nouvelles connaissances portent sur la théorie explicative des impacts.

Quelle que soit la finalité affichée, une évaluation réussie doit obligatoirement avoir une utilisation cognitive visible et reconnue.

● Utilisation pour décider des améliorations à apporter

▶ L'équipe d'évaluation et le groupe de pilotage formulent des recommandations aux décideurs. La participation active du groupe de

pilotage à cette étape de l'évaluation a l'avantage de donner aux recommandations une dimension plus réaliste et de tenir compte des préoccupations des protagonistes qui pourront par la suite se les approprier plus facilement. Quel que soit le choix fait, il est important que les recommandations soient nettement séparées des conclusions et que leur statut soit clairement expliqué en répondant aux questions suivantes : la mise en œuvre de cette recommandation a-t-elle une implication financière ? peut-elle intervenir à court terme ? la recommandation fait-elle consensus ? Même s'il n'existe jamais d'obligation de se conformer aux recommandations d'une évaluation, il est intéressant que le commanditaire suive pendant un ou deux ans le devenir de ces recommandations.

Dans le meilleur des cas, il est possible de retrouver la trace des recommandations dans les décisions qui ont été prises à la suite de l'évaluation. Celle-ci constitue alors un instrument d'aide à la décision et peut servir à améliorer le programme en cours de route ou à l'occasion de son renouvellement, comme cela a été le cas dans la situation présentée ci-dessous. Cette utilisation des conclusions de l'évaluation est qualifiée « d'instrumentale ».

Exemple : l'utilisation instrumentale de l'évaluation

Le rapport de l'évaluation intermédiaire du CCA irlandais 1994-99 a représenté une importante base de discussions lors du bilan à mi-parcours dudit CCA. Suite aux négociations intervenues entre la Commission européenne et les autorités nationales, le Comité de suivi s'est mis d'accord sur toute une série d'ajustements portant sur près de 3% du montant global. Une partie des principales ré-affectations financières résulte clairement des recommandations issues de l'évaluation du CCA, tout comme certaines décisions autres que financières. Ainsi, sur les dix huit mesures ou groupe de mesures pour lesquels les allocations étaient remises en question, huit ont vu cette aide diminuer. Par contre parmi les huit domaines qui avaient été identifiés comme souffrant de besoins d'aides complémentaires, cinq se sont vus attribuer une aide additionnelle. Des décisions non financières ont été prises dans le sens d'un changement de priorités : formation et renforcement des capacités d'expansion pour les petites entreprises au détriment de l'aide au secteur privé en général.

● Utilisation pour juger les succès et les échecs

► Les utilisateurs trouvent dans l'évaluation des arguments pour former ou modifier leur jugement. En effet, l'évaluation va aider à expliciter des critères et à définir des normes de comparaison. Elle va conduire à tirer des

conclusions sur le succès ou l'échec de telle ou telle partie du programme et par conséquent à identifier des bonnes pratiques et des projets exemplaires.

Cette forme d'utilisation est primordiale lorsque l'évaluation sert à rendre compte aux autorités politiques et au grand public. Il s'agit d'une utilisation externe au programme qui alimente le débat sur la justification de ce programme. L'évaluation s'apparente alors à un compte-rendu et est destinée à permettre à des personnes étrangères au contexte du programme (les pouvoirs publics, les élus, le grand public) de se former une opinion globale sur la valeur intrinsèque de l'action, indépendamment de l'opinion des protagonistes.

Toujours dans ce contexte, l'utilisateur confronte les conclusions de l'évaluation à ses propres opinions. Il en vient à maintenir, conforter ou changer son jugement personnel sur le programme. Cette autre forme d'utilisation est appelée normative en raison du fait qu'elle comporte une forte part de jugement.

Toutes les évaluations ne parviennent pas à cette dernière forme d'utilisation car l'utilisation normative est la plus délicate. En effet, si un utilisateur doit réviser son jugement, il risque de manifester une certaine résistance, surtout si son jugement a déjà été exprimé en public auparavant.

La littérature spécialisée a coutume de distinguer deux pratiques d'évaluation selon qu'elle est principalement destinée aux acteurs directement concernés par le programme ou que les destinataires sont exogènes. L'évaluation est appelée endoformative ou « formative » lorsque son utilisation, interne à l'action ou au programme évalué, est de nature à donner aux protagonistes du programme les informations qui leur permettront de modifier leurs conduites, d'améliorer l'action et de transformer ainsi son objet même.

L'évaluation est dite récapitulative quand elle sert à la fois à juger le programme et à aider les autorités de tutelle à décider de son lancement, de sa reconduction ou de sa suppression. Les utilisateurs sont alors les financeurs extérieurs et éventuellement d'autres acteurs sociaux qui peuvent décider de l'opportunité d'adopter pour eux-mêmes les meilleures pratiques identifiées à l'occasion de l'évaluation du programme.

Quelle que soit la nature de la décision qui est prise à la suite de l'évaluation, le processus d'utilisation passe successivement par les trois

formes d'utilisation. Le « décideur » commence par transformer sa façon de comprendre le programme, puis il formule, renforce ou modifie son jugement, avant de s'exprimer au bon moment et de façon convaincante pour entraîner ses co-décideurs ; la décision publique étant presque toujours prise à plusieurs.

Pour passer de l'évaluation à la décision, on voit que les étapes logiques sont nombreuses et que les dernières étapes mettent en jeu plusieurs acteurs. On comprend donc que l'utilisation instrumentale d'une évaluation soit difficile à démontrer. C'est ce que confirment les travaux de méta-évaluation qui concluent souvent à l'absence de lien direct à court terme entre les recommandations et les décisions.

◀ C'est en ce sens que l'évaluation se distingue d'une étude de faisabilité du programme. En effet, cette dernière est toujours réalisée sur la base des objectifs du responsable politique pour préparer directement sa décision et ne comporte pas de remise en question de ses objectifs. Elle a pour seule et unique vocation de déboucher sur des décisions. Au contraire, une évaluation comporte toujours une dimension de jugement sur l'opportunité du programme en prenant en compte les objectifs politiques d'une part et, d'autre part, le point de vue des destinataires. Même dans le cas d'une évaluation *ex ante* qui porte sur une proposition de programme, l'utilisation instrumentale n'est pas la seule et unique modalité d'utilisation ; les dimensions cognitive et normative sont inévitablement présentes.

Par ailleurs l'évaluation diffère aussi d'un travail de recherche fondamentale du point de vue de son utilité car elle répond à une commande des responsables du programme, à un moment où ils sont demandeurs d'information et de réponses aux questions qu'ils se posent. L'exercice est toujours sous des contraintes fortes que sont les hypothèses d'action et le champ d'investigation choisi par les décideurs. En contrepartie, les résultats des évaluations sont souvent mieux acceptés par les pouvoirs publics, et donc davantage utilisés.

Apprécier la qualité d'une évaluation

La qualité d'une évaluation est souvent jugée sur la seule base du rapport final, or elle dépend très largement de la façon dont on a su gérer la qualité de l'ensemble du processus depuis la commande jusqu'à la formulation de recommandations et à la communication des résultats.

Juger de façon équitable la qualité d'une évaluation suppose que l'on explicite les critères utilisés. L'appréciation peut s'appuyer sur huit critères qui constituent les standards professionnels importants à connaître : la satisfaction des demandes, la pertinence du champ, l'adéquation de la méthodologie, la fiabilité de ses données, la solidité de l'analyse, la crédibilité des résultats, l'impartialité des conclusions et la clarté du rapport.

Il convient d'apprécier la qualité de l'évaluation pour vérifier si l'équipe a satisfait les exigences des commanditaires mais également pour distinguer, parmi les conclusions, celles qui sont valides et solidement étayées de celles qui sont susceptibles d'être contestées ou qui nécessitent une vérification empirique plus approfondie.

Apprécier la qualité de l'évaluation est, ou devrait être, un préalable à l'utilisation de ses conclusions et recommandations. C'est un préalable logique car seule une évaluation de bonne qualité peut résister aux critiques que suscitent inévitablement les jugements portés sur les succès et les échecs. La qualité est aussi un préalable chronologique car elle fait le plus souvent l'objet d'une appréciation formelle avant la diffusion du rapport d'évaluation et avant la prise en considération de ses conclusions et recommandations.

Nous avons choisi toutefois de placer ce chapitre à la fin des trois parties du Volume qui sont consacrées à la conception d'un projet d'évaluation et sa réalisation car la grille d'appréciation de qualité représente une checklist de l'ensemble des propositions et recommandations faites dans les chapitres précédents. Elle a été mise au point dans le cadre du programme MEANS avec l'aide d'experts de plusieurs pays européens et pourra donc servir au lecteur qui le souhaiterait pour vérifier si sa propre pratique répond aux standards professionnels courants.

■ Gérer et améliorer la qualité d'un processus d'évaluation

Avec le développement de l'évaluation et par conséquent l'augmentation générale des budgets consacrés à cette activité, l'attention se porte de plus

en plus sur la qualité des travaux rendus, mais aussi sur la qualité de l'ensemble du processus qui va de la commande jusqu'à l'utilisation de ces travaux.

L'appréciation de la qualité vérifie que le contrat de l'équipe d'évaluation est rempli. Son existence même contribue à sensibiliser l'équipe d'évaluation, ce qui conduit à améliorer le rapport final. Plus indirectement, l'appréciation peut améliorer les pratiques professionnelles en diffusant des standards de bonne pratique. Elle peut aussi servir un objectif démocratique plus large en renforçant la crédibilité des informations diffusées par les administrations publiques.

Les critères de qualité qui seront présentés plus loin ont été utilisés par la Commission pour l'appréciation d'un certain nombre de rapports d'évaluation à mi-parcours de la période de programmation 1994-1999. Ils ont également été appliqués par certaines autorités nationales. Ainsi par exemple, la qualité de plusieurs dizaines d'évaluations réalisées en République d'Irlande au cours de la période 1994-1998 a été appréciée en se référant à ces critères. Ces travaux s'apparentent à des exercices de méta-évaluation, c'est à dire à l'évaluation d'une ou plusieurs évaluations.

La qualité est une exigence suffisamment cruciale pour qu'elle devienne aussi la préoccupation de chaque commanditaire d'évaluation. L'appréciation de la qualité relève notamment des co-décideurs de chaque programme, sous forme d'auto-appréciation pour ce qui concerne la qualité de la commande passée, et sous forme de contrôle de qualité pour ce qui concerne le travail effectué par l'équipe d'évaluation.

Les co-décideurs ne sont cependant pas les seuls acteurs à trouver un intérêt dans la qualité des rapports (voir Encadré 44). En particulier, le groupe de pilotage décide si la collecte et l'analyse des données sont suffisantes ou doivent être améliorées, si les conclusions sont recevables au vu des éléments de preuves recueillis, si le rapport répond à toutes les questions du cahier des charges ou s'il doit être complété.

En général, les autorités régionales, nationales et européennes analysent elles-mêmes la qualité des évaluations en fonction de leurs propres contraintes et critères. Dans le cas d'évaluations de haute technicité ou de diffusion publique de conclusions sensibles, le commanditaire peut faire appel à un spécialiste extérieur qui se prononcera sur la qualité du processus et qui effectuera une contre-expertise de la robustesse des

conclusions. Les critères de qualité présentés plus loin sont en grande partie issus des standards professionnels, développés dans le cadre d'expertises de ce type.

Encadré 44 - Qui apprécie la qualité des évaluations ?

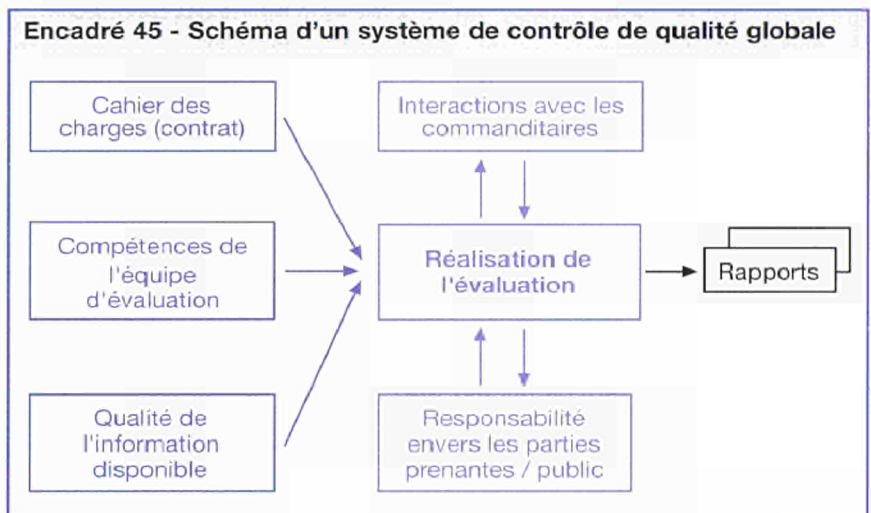
<i>L'appréciation de qualité peut être faite par</i>	<i>en vue de</i>
le groupe de pilotage avec le cas échéant l'aide d'un spécialiste	de valider le rapport final et vérifier sa robustesse
les co-décideurs du programme	de connaître la fiabilité des conclusions et des recommandations
les autorités nationales et européennes	- de connaître la fiabilité des conclusions et des recommandations - ou d'améliorer le système d'évaluation en vigueur
des experts indépendants	- de garantir la sincérité des conclusions diffusées au public - développer des standards professionnels (méta-évaluation)

L'appréciation peut être conduite dans un esprit d'assurance de la qualité de l'ensemble du processus d'évaluation. En effet, le rapport final n'est que la résultante d'un système qui mobilise plusieurs personnes au sein d'un vaste ensemble institutionnel et d'un dispositif spécialement mis en place pour l'évaluation proprement dite. Dans ce cas, il s'agit de surveiller et de promouvoir la qualité aux grandes étapes de l'évaluation. Si l'on considère le déroulement de l'évaluation depuis sa commande jusqu'à la production du rapport et la diffusion des résultats, on peut améliorer la qualité d'une évaluation en agissant sur plusieurs points, à savoir sur :

- La qualité du système d'information et de suivi.
- Les termes du cahier des charges qui définissent ce qui doit être évalué, les ressources budgétaires allouées et le temps imparti.
- Le choix de l'équipe d'évaluation (qualifications, références, expériences, etc.).
- Le contrat passé avec l'équipe d'évaluation, qui définit notamment son plan de travail.

- La fréquence et la nature des interactions entre l'équipe d'évaluation et le groupe de pilotage.
- La fourniture rapide et complète des informations disponibles de qualité.
- Le mode de communication des conclusions comprenant le rapport lui-même et les autres canaux de diffusion.
- L'utilisation de l'évaluation pour rendre des comptes aux autorités publiques et aux citoyens.

Un système d'assurance de la qualité porte sur l'ensemble du processus d'évaluation et intervient donc sur tous ces points. Il peut être schématisé comme le montre l'encadré ci-dessous.



■ Analyser la qualité du rapport d'évaluation

Dans les textes généraux consacrés à l'évaluation, la notion de qualité fait l'objet de nombreux débats. Il n'existe dans le monde aucun système de certification professionnelle institutionnalisant des règles et des critères de qualité. Des grilles de critères assez disparates sont proposées en fonction des modèles d'évaluation élaborés par les différents auteurs mais il n'existe pas de consensus dans ce domaine et la nature des critères mentionnés n'apparaît pas toujours clairement lorsque l'on essaie de les mettre en pratique.

Dans le cadre de MEANS, huit critères d'appréciation de la qualité des rapports d'évaluation ont été retenus. L'appréciation globale d'une évaluation est établie de façon synthétique après que le rapport final ait été analysé au regard de ces huit critères. Ils sont présentés ci-dessous en prenant comme situation de référence les évaluations à mi-parcours. La grille a cependant une portée plus générale et peut être appliquée à toutes les situations d'évaluation sachant que la formulation de certaines questions et le poids relatif des critères doit tenir compte de la spécificité de l'exercice considéré. Dans la suite du texte nous avons mentionné de telles adaptations pour l'évaluation ex ante et l'évaluation ex post par rapport à l'évaluation intermédiaire lorsqu'elles existent.

Huit critères ont été définis pour apprécier la qualité des rapports d'évaluation. L'appréciation globale d'une évaluation est établie de façon synthétique après que le rapport final ait été analysé au regard de ces huit critères.

● La satisfaction de la demande

Dans le cadre des programmes des Fonds structurels européens, les principaux protagonistes sont les membres du groupe de pilotage. Dans l'idéal, leurs diverses demandes sont réunies dans le cahier des charges. Cependant, les questions évaluatives qui y sont formulées doivent être complétées lors de réunions et d'entretiens préalables afin d'aller au-delà de ce qui est déjà évident pour les commanditaires. Par conséquent, on vérifiera, d'une part, si l'évaluation traite de manière adéquate les questions du cahier des charges et, d'autre part, si le rapport traite les demandes supplémentaires d'information, ainsi que des questions nouvelles qui seraient essentielles pour les commanditaires.

Par conséquent, pour savoir si « l'évaluation traite de façon adéquate les demandes d'information formulées par les commanditaires et correspond au cahier des charges » (Cf. la grille synthétique de l'Encadré 46 page 192) on se posera plus particulièrement les questions suivantes :

- L'évolution des objectifs globaux et spécifiques du programme a-t-elle été analysée ?
- Le rapport aborde-t-il la totalité du programme ? Si tel n'est pas le cas, la sélection faite est-elle justifiée au regard des priorités exprimées par les commanditaires ?
- L'évaluation apporte-t-elle des informations en retour utiles pour les gestionnaires ?

- Y trouve-t-on des enseignements sur les échecs et les réussites qui puissent intéresser d'autres régions ou pays ?

Pour les évaluations ex post, il sera important de vérifier si les informations nécessaires aux prochaines générations de programmes ont été pris en considération. Aussi, devra-t-on regarder si l'évaluation a réussi à trouver un bon compromis entre les deux exigences contradictoires que sont : obtenir rapidement des informations pour alimenter les nouveaux cycles de programmation d'une part, et ne pas tirer de conclusions précipitées avant que l'ensemble des impacts soit observable, d'autre part.

● La pertinence du champ

Le champ de l'évaluation doit couvrir les questions qui sont pertinentes du point de vue des programmes des Fonds structurels. La description succincte des caractéristiques essentielles du programme et l'analyse de sa stratégie pour la région concernée sont des préalables indispensables pour la compréhension de l'action évaluée et de sa raison d'être.

Afin de vérifier la pertinence du champ, on regardera si, premièrement, les caractéristiques essentielles du programme sont bien décrites (raison d'être, stratégie, objectifs, actions, ressources, réalisations etc.) et, deuxièmement, si les obstacles et les succès dans la mise en œuvre du programme sont bien explicités.

De plus, les réalisations, les résultats et les impacts du programme devant être couverts par l'analyse pour juger de l'atteinte des objectifs du programme, on vérifiera comment l'évaluation les inclue. De surcroît, on s'assurera que l'évaluation ne néglige pas d'autres résultats ou impacts potentiels ou futurs ainsi que d'éventuels résultats et effets inattendus et significatifs.

Le champ de l'évaluation dépend enfin de la cible du programme qui peut être définie en termes de zones géographiques éligibles ou de groupes cibles non localisés (les chômeurs de longue durée, etc.). Par conséquent il s'agit de vérifier si :

- les limites du champ, en termes de zones ou de groupes sociaux, sont définies de manière appropriée eu égard à la logique de l'intervention ;
- le champ inclut des zones périphériques ou des groupes non-éligibles mais susceptibles d'être affectés par les interventions évaluées ;
- et encore, si l'évaluation considère le programme évalué de façon isolée ou inclut ses interactions avec d'autres programmes européens et nationaux.

Dans le cas des évaluations ex ante, le rapport doit ajouter toutes les informations nouvelles utiles à l'amélioration des documents de programmation : stratégie, cohérence interne/externe, mise en œuvre des choix. On se demandera si le rapport évalue avec suffisamment de détails les attentes des groupes cibles ainsi que la pertinence des objectifs et de la stratégie.

● L'adéquation de la méthode

Ce critère essentiel a trait aux qualités techniques de l'évaluation. Les choix de méthodes doivent découler des questions évaluatives et des démonstrations requises. L'évaluation doit par ailleurs, faire le meilleur usage des recherches et des études existantes. L'application de ce critère (ainsi que des deux suivants) nécessite des connaissances techniques. Le cas échéant, les commanditaires pourront, prendre l'avis d'un spécialiste de l'évaluation à ce stade de leur appréciation.

Dans tous les cas, trois types de question devront être posées :

- Les connaissances pertinentes et la littérature existante ont-elles été collectées et utilisées à bon escient ?
- La construction de la méthode et le choix des outils sont-ils véritablement justifiés pour répondre au mieux aux questions évaluatives ?
- Les situations de référence (contrefactuelles ou similaires) choisies étaient-elles adaptées pour faire des comparaisons valables ?

Tout rapport d'évaluation doit inclure une description de la méthode utilisée et définir clairement les sources de données. De même, les limites de la méthode construite et des outils utilisés doivent être clairement exposées. On regardera si :

- la méthode est décrite de façon suffisamment détaillée pour permettre d'en juger la qualité ;
- la validité des données collectées et des outils utilisés est clairement indiquée ;
- et enfin, si il y a adéquation entre les données disponibles et les outils utilisés.

Pour les évaluations ex post, l'analyse causale des effets étant la question la plus importante, les méthodes d'analyse de ces relations causales sont prioritaires dans ce type d'évaluation. On regardera dans quelle mesure

l'évaluation analyse suffisamment les relations de cause à effet pour les questions les plus essentielles.

● **La fiabilité des données**

Les chargés d'évaluation utiliseront des données existantes (données secondaires) issues du système de suivi et d'autres sources d'information ou bien des données primaires qu'ils recueilleront. Dans ce dernier cas, le mode de collecte des données (choix et application des outils utilisés à cette fin) ainsi que le traitement de celles-ci seront très importants pour la fiabilité et la validité des résultats.

Afin d'apprécier la fiabilité des données utilisées, on examinera :

- si toutes les sources d'information qui étaient disponibles ont été identifiées et si la fiabilité de ces données a été vérifiée ;
- si les sources d'informations issues du système de suivi et des études précédentes ont été utilisées de façon optimale ;
- si les techniques de collecte des données choisies étaient complètes et adaptées pour répondre aux questions évaluatives ;

Que l'équipe d'évaluation ait eu recours à des techniques quantitatives ou qualitatives ou encore à une combinaison des deux, la qualité technique des données collectées doit être vérifiée : définition des « populations-mères », sélection et échantillonnage, construction des instruments (le questionnaire, les indicateurs choisis), etc. On devra alors se demander si :

- le mélange de données qualitatives et quantitatives est adapté pour faire une analyse valide du phénomène ;
- les « populations » utilisées pour la collecte des données sont correctement définies ;
- la sélection des échantillons d'enquête ou des cas étudiés est effectuée en fonction de critères justifiés ;
- et si les principales techniques de collecte de données sont mises en œuvre avec leurs propres normes de qualité de manière à garantir un degré suffisant de fiabilité et de validité des résultats.

Si l'évaluation se limite aux données secondaires à cause des contraintes de temps (cas fréquent de l'évaluation ex ante), cela suppose que les sources d'information sélectionnées soient suffisamment fiables et généralisables.

● La solidité de l'analyse

L'analyse quantitative se traduit par l'analyse des données sous forme de tableaux ou de tout autre type d'analyse statistique. L'analyse qualitative comporte une comparaison et un recoupement systématique des sources d'information. Quel que soit le cas, il est nécessaire d'apprécier si les méthodes d'analyse choisies sont pertinentes au regard du type de données recueillies et, si l'analyse a été menée conformément aux prescriptions qui figurent dans les manuels techniques.

Dans le cas des Fonds structurels, les relations de cause à effet sont complexes et constituent par conséquent un défi particulier pour l'évaluation. On devra vérifier :

- si l'évaluation explicite suffisamment les relations de cause à effet qui sont sous-jacentes au programme,
- et dans quelle mesure l'analyse s'appuie sur des techniques adaptées.

Dans le cadre d'une évaluation plus approfondie il s'agit de vérifier s'il y a eu une comparaison avant/après impliquant des destinataires et un groupe de comparaison. Il faut que l'analyse comparative se soit appuyée sur une collecte des données au sein du groupe de comparaison.

Par ailleurs, l'exploitation des travaux d'évaluation réalisés au niveau des programmes doit permettre à la Commission de préparer des synthèses au niveau européen. Cela suppose qu'il y ait une cohérence des définitions (par exemple : calcul de l'effet net, etc.), une homogénéité des chiffres, etc. On vérifiera si les résultats de l'évaluation permettent des comparaisons de performances avec d'autres programmes similaires.

● La crédibilité des résultats

La crédibilité est ici définie comme le fait que les résultats découlent logiquement des données collectées et analysées et que les interprétations sont basées sur des hypothèses explicatives présentées avec soin. Dans le cas d'une analyse approfondie, la qualité de l'évaluation tient compte du compromis à trouver entre la validité interne (prise en compte complète et sans biais de toutes les influences contextuelles) et la validité externe (généralisation possible à d'autres contextes). Lorsque l'évaluation comporte des analyses approfondies sur une partie du programme (par exemple des études de cas de quelques actions) et que les résultats de ces analyses sont extrapolés à tout ou partie du programme, on devra vérifier :

- si les hypothèses interprétatives et les extrapolations faites sont justifiables et si les limites de validité ont été exposées ;
- et si la sélection des cas et des échantillons permet une généralisation.

La participation des parties prenantes à la discussion des résultats permet d'accroître la pertinence de l'évaluation et donc son utilité pour le programme. On se posera les questions suivantes :

- Les résultats sont-ils crédibles dans le sens où ils reflètent un compromis acceptable entre la réalité du programme telle qu'elle est perçue par les acteurs et la réalité décrite par les chiffres et les faits constatés ou estimés ?
- Dans quelle mesure les parties prenantes ont-elles été impliquées dans le processus d'interprétation des observations empiriques ?
- Les résultats ont-ils été présentés de façon impartiale aux parties prenantes ?

● **L'impartialité des conclusions**

Les conclusions comprennent les suggestions et parfois même des recommandations qui vont au-delà des simples résultats. Alors que les résultats sont « techniques » et peuvent être analysés sans trop de risque de partialité, les conclusions et *a fortiori* encore plus les recommandations, sont, quant à elles, émises sur la base de jugements de valeur. Aussi, la qualité du jugement est-elle déterminante.

Pour répondre à la question : les conclusions sont-elles justes, exemptes de considérations personnelles ou partisans, et sont-elles assez détaillées pour être concrètement mises en oeuvre ? on vérifiera si :

- les éléments sur lesquels sont basées les conclusions sont bien explicités ;
- les conclusions sont opérationnelles et suffisamment explicitées pour être mises en oeuvre ;
- et si les questions controversées sont présentées de façon équilibrée et équitable.

● **La clarté du rapport**

La diffusion et la communication des résultats de l'évaluation auprès des parties prenantes peuvent se faire par voie écrite ou par voie orale. Le rapport final n'est qu'un moyen de diffusion parmi d'autres ; une

communication continue des résultats est même souhaitable. La lisibilité du rapport et la qualité de sa présentation sont des éléments décisifs de la qualité de l'évaluation. On regardera :

- si le rapport a été rédigé de façon claire pour ses destinataires et si sa structure est logique ;
- si les concepts spécialisés sont utilisés uniquement lorsque cela est indispensable et s'ils sont clairement définis ;
- dans quelle mesure la présentation, les tableaux et les graphiques renforcent la lisibilité et l'intelligibilité du rapport ;
- et si les limites de l'évaluation, en termes de portée, de méthodes et de conclusions sont clairement exposées.

La lecture d'un rapport se limite souvent à la lecture de son résumé. Par conséquent, il est essentiel que celui-ci soit clair et concis. Il doit présenter les principales conclusions et recommandations de façon équilibrée et impartiale. Il doit se lire aisément sans nécessiter la lecture complète du rapport.

On s'attachera aussi au fait que les conclusions soient claires pour les lecteurs intéressés par telle ou telle priorité européenne ou nationale, par exemple l'impact environnemental, l'impact sur l'emploi, l'égalité des chances, etc.

■ Porter une appréciation globale et équitable

L'appréciation synthétique récapitule tous les critères de qualité. Il n'y a pas lieu de recommander une pondération particulière des différents critères car leur importance varie d'une situation à l'autre. L'appréciation synthétique est réalisée avec l'aide de la grille synthétique figurant dans l'Encadré 46.

L'appréciation synthétique doit tenir compte des contraintes qui pèsent sur l'évaluation et l'équipe qui l'a réalisée. Il faut donc nuancer la première conclusion à laquelle on est parvenu en appliquant strictement les critères de qualité. Par exemple, un rapport peut être jugé peu adéquat non en raison de déficiences liées aux travaux proprement dits mais parce que le cahier des charges était lui-même irréaliste. Il est fréquent de constater que les ressources allouées et le temps imparti à l'équipe d'évaluation limitent excessivement la portée et le degré de robustesse des conclusions. Les contraintes contextuelles doivent donc être prises en considération pour juger *in fine* le rapport.

On peut distinguer deux types de contraintes selon qu'elles relèvent de la commande de l'évaluation ou de la gestion du programme évalué. Les contraintes relatives à la procédure de commande sont la clarté et la précision du cahier des charges, le volume des ressources budgétaires, le temps alloué et l'implication des commanditaires pendant la mise en œuvre de l'évaluation. Les contraintes relatives au programme et à sa gestion peuvent résulter de défauts de clarté des objectifs, d'une forte hétérogénéité entre les mesures et les interventions, d'une qualité insuffisante des données générées par le système de suivi, ainsi que de crises ou ruptures qui peuvent intervenir dans la mise en œuvre et qui ont des conséquences sur le déroulement de l'évaluation.

Il incombe aux commanditaires de limiter, si ce n'est de faire disparaître, les conséquences de ces contraintes, mais aussi aux équipes d'évaluation d'anticiper les problèmes techniques et opérationnels avant même de s'engager dans le processus d'évaluation.



V

Assurer la réussite d'une évaluation



Mieux préparer l'évaluation

Les programmes financés par les Fonds structurels font l'objet d'un effort d'évaluation particulièrement intense. Depuis 1989, plusieurs centaines de programmes ont été évalués dans leur globalité, avant, pendant et après leur mise en œuvre, sans compter les évaluations thématiques transnationales et les analyses plus approfondies qui sont réalisées dans certaines régions. Il existe peu de cas où une pratique d'évaluation se soit répandue de façon aussi rapide et aussi massive. Des progrès restent cependant réalisables tant du point de vue du rapport qualité/coût que du point de vue de l'utilisation.

Parmi les recommandations susceptibles de renforcer l'utilisation de l'évaluation, un certain nombre concernent sa préparation. La planification pluriannuelle des différents exercices d'évaluation doit se poursuivre. Par ailleurs il conviendrait que les autorités commanditaires accordent une plus grande visibilité à leur décision quand elles lancent une évaluation et qu'un investissement préalable plus important soit effectué pour la rédaction des cahiers des charges, au besoin sous forme d'une étude d'évaluabilité lorsque des questions complexes sont à traiter.

■ Faire une programmation pluriannuelle des évaluations

Comme cela a été décrit dans la deuxième Partie de l'ouvrage (voir Encadré 7), la succession des programmes, période après période, est une donnée incontournable. Les programmes se suivent sans interruption de calendrier et le cadre réglementaire impose une succession d'évaluations à chaque période (ex ante / à mi-parcours / ex post). La mise en pratique de ces différentes contraintes a créé et continuera de créer des problèmes d'articulation entre les différents exercices d'évaluation. En particulier, il s'avère souvent difficile de dégager les leçons de la mise en œuvre d'un programme à un moment et dans une forme qui soit réellement adaptée aux enjeux décisionnels. L'expérience montre qu'il est nécessaire de réfléchir à un programme pluriannuel d'évaluation dans le but de renforcer des opportunités pour la rétroaction (processus qui conduit de l'observation des résultats et des impacts jusqu'aux décisions de réorientation du programme).

Afin de traiter cette difficulté, nombreux sont les gestionnaires et les équipes d'évaluation qui souhaitent le développement d'une évaluation en

continu (*on-going evaluation, rolling evaluation*). Des expériences en la matière ont été conduites en Espagne (pour les programmes opérationnels régionaux d'Andalousie et de la Communauté Autonome de Valence), en Irlande (pour environ la moitié des programmes opérationnels), en France (programmes de la région Bourgogne), etc. Ces tentatives méritent d'être étudiées avec attention.

De nombreux experts sont cependant dubitatifs par rapport à l'évaluation conçue comme processus permanent. Ce système crée en effet un risque de confusion entre l'évaluation, le suivi et la gestion. Il peut en résulter une dilution de la valeur ajoutée de l'évaluation. En revanche, il est possible de recommander la mise en place d'un programme pluriannuel d'évaluation articulant une succession des travaux de différente nature à différents stades : évaluation *ex ante*, analyses approfondies, évaluation globale à mi-parcours, évaluation *ex post*. Cette recommandation s'inspire d'expériences australiennes et canadiennes. Elle a été reprise par la Commission dans le cadre de son propre système d'évaluation interne (SEM 2000).

Un tel programme permettrait d'articuler les différents exercices entre eux et de bénéficier à tout moment des leçons tirées de l'expérience passée, que ces leçons aient émergé d'une période de programmation ou d'une autre. De plus, un tel programme permettrait de coordonner les travaux d'évaluation commandés à différents niveaux sur un même territoire et d'éviter d'éventuelles répétitions d'enquêtes auprès des mêmes personnes.

Il existe plusieurs exemples où les commanditaires ont jugé préférable de faire appel à une seule équipe pour toute une succession de travaux d'évaluation dans une même région ou dans un même domaine d'intervention. Cette formule a l'avantage d'assurer une coordination automatique des évaluations qui sont toutes réalisées par le même prestataire. Elle ne peut cependant pas être prolongée indéfiniment sous peine de compromettre l'indépendance des évaluations.

■ Impliquer les décideurs à travers un mandat d'évaluation

Même si l'évaluation est une obligation réglementaire, la décision d'évaluer (voir page 67) reste un acte politique qui résulte d'une décision du commanditaire principal. De nombreuses expériences ont montré que cette décision gagne à être formalisée dans un document bref, clair et visible, qui est souvent appelé « mandat d'évaluation ». Ce document prend toute sa

valeur s'il est personnellement signé par la plus haute autorité hiérarchique concernée par l'évaluation.

Le mandat précise ce qui va être évalué (le champ de l'évaluation), les motifs pour lesquels l'évaluation est lancée et l'utilisation qui en est attendue. Le commanditaire principal précise s'il attend des recommandations à l'issue de l'évaluation ou s'il n'en attend pas. Il énonce quelques grandes questions auxquelles l'évaluation devra répondre, ce qui donne généralement une première idée des critères d'évaluation. Il fixe une date limite pour la remise du rapport et éventuellement une enveloppe budgétaire.

En procédant de cette façon, le commanditaire fixe les limites qui lui paraissent raisonnables et qui s'appliqueront à l'ensemble du processus d'évaluation. Le mandat évite cependant de trop encadrer la liberté du groupe de pilotage et de l'équipe d'évaluation. Il ne doit pas devenir un instrument de censure préventive. Son principal avantage est au contraire de concentrer l'évaluation sur des sujets stratégiques, intéressants au plus haut niveau, et susceptibles de favoriser l'utilisation des conclusions. Le mandat évite aussi à l'évaluation de s'installer sur des terrains politiquement minés au risque d'être totalement rejetée au moment de la sortie des résultats.

Le mandat vise à prévenir un certain nombre de dérives qui peuvent être énumérées de la façon suivante :

- Les questions stratégiques sont perdues de vue et les préoccupations à court terme des gestionnaires l'emportent.
- Les autorités ne se sentent pas engagées par les conclusions, les responsables hiérarchiques n'autorisent pas leurs subordonnés à participer à l'évaluation.
- L'équipe d'évaluation dépasse son rôle.
- Le rapport d'évaluation arrive trop tard et son utilisation est impossible.
- L'évaluation a tendance à trop embrasser et ne peut pas arriver à conclure.

Un autre avantage du mandat est d'établir des règles du jeu claires et réalistes. Dans cette perspective, l'autorité signataire confie officiellement la responsabilité du processus d'évaluation à une personne, assistée de préférence par un groupe de pilotage. La décision fixe la composition du groupe de pilotage et les limites de ses missions.

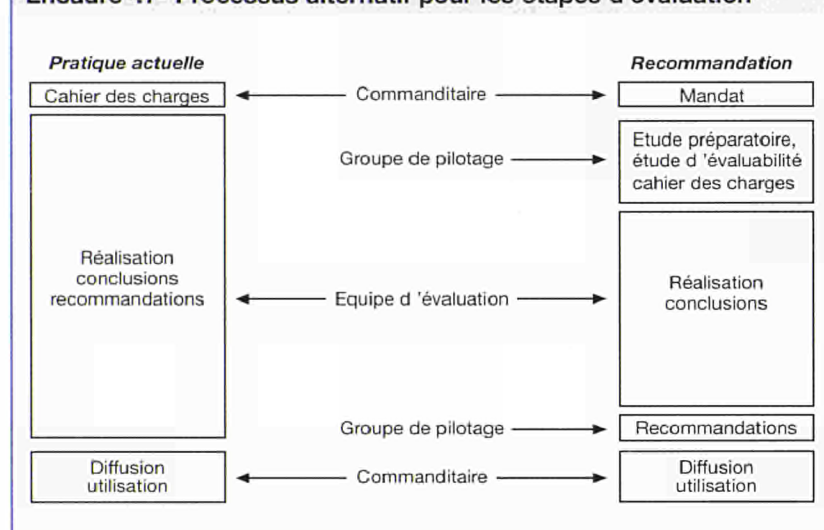
■ Investir dans la préparation du cahier des charges

L'expérience a montré que le cahier des charges est souvent rédigé rapidement sur la base de modèles ou de documents types. Cette pratique risque de donner à l'évaluation un caractère bureaucratique, peu orienté vers la spécificité régionale, peu adapté aux enjeux du programme et peu motivant pour les gestionnaires. Aussi longtemps que la rédaction du cahier des charges est considérée comme un travail de bureau, il est difficile de faire les choix nécessaires pour qu'elle soit véritablement adaptée aux besoins de ses utilisateurs futurs.

La situation est différente si le commanditaire a établi dès le départ un groupe de pilotage officiellement mandaté pour toute la durée de l'évaluation. On peut alors faire jouer à ce groupe un rôle actif dans la préparation du cahier des charges (voir page 79). Cela suppose d'organiser une ou deux réunions, ou parfois trois, qui vont surtout permettre de formuler les questions évaluatives sans ambiguïté, de vérifier que des questions essentielles (ou embarrassantes) n'ont pas été oubliées, d'écarter des questions qui relèveraient plutôt de l'audit ou du contrôle, de sélectionner les questions prioritaires, de préciser les critères et les normes de jugement. L'expérience a montré qu'une telle étude préparatoire conduit à envisager plusieurs stratégies d'évaluation différentes, oblige à faire des choix et améliore finalement la qualité de l'évaluation. Un choix extrême, mais salutaire, peut d'ailleurs être celui du report ou de l'abandon de l'évaluation lorsque le groupe de pilotage anticipe qu'elle ne pourra pas être utilisée.

L'étude préparatoire ne constitue pas un travail supplémentaire mais plutôt une organisation alternative de la rédaction du cahier des charges de l'évaluation. Les deux réunions que le groupe de pilotage y consacre permettent de réaliser ce qui est habituellement fait par l'équipe d'évaluation, dans le cadre de la mise en place de ses tâches, comme le montre l'Encadré 47.

Il s'agit de revenir à la raison d'être du programme et à sa stratégie, de récapituler ses objectifs, d'afficher la hiérarchie et l'articulation de ces objectifs. A ce stade, le groupe de pilotage vérifie qu'il n'y a pas de malentendu à propos de la logique du programme, et il apporte toute clarification qui serait nécessaire. Au cours des réunions préparatoires, le groupe de pilotage rédige avec soin les questions évaluatives et sélectionne les plus importantes. Il se prononce sur les critères de jugement et les normes de comparaison qui seront appliqués par l'équipe d'évaluation.

Encadré 47- Processus alternatif pour les étapes d'évaluation

Dans le cadre de l'étude préparatoire, les responsables du programme réalisent par eux-mêmes des tâches qu'ils sont les seuls à pouvoir faire de façon compétente et légitime. Ils évitent la situation fréquente où l'équipe d'évaluation reproduit, avec une perte de qualité, des informations déjà connues de l'administration. Au contraire, l'équipe d'évaluation peut concentrer son travail sur la production de données et d'analyses nouvelles.

L'étude d'évaluabilité est un exercice voisin qui peut être éventuellement combiné avec les réunions préparatoires. Il s'agit d'une vérification technique qui est recommandée par plusieurs auteurs nord-américains. L'étude porte sur l'évaluation dans son ensemble et sur chaque question en particulier. Elle fait le point sur le niveau des connaissances disponibles et conclut sur le fait que les conditions techniques et institutionnelles seront réunies pour apporter des réponses fiables et crédibles aux questions posées. Ce diagnostic est fait en peu de temps par un méthodologue de l'évaluation (membre d'une unité d'évaluation ou expert extérieur). Il conduit à abandonner certaines questions considérées comme trop difficiles, ou même des questions pour lesquelles on s'aperçoit qu'une réponse existe déjà.

L'étude d'évaluabilité définit les grandes lignes de la méthode d'évaluation, ce qui permet de diriger l'appel d'offre vers des équipes disposant des compétences nécessaires. Elle permet également d'ajuster le nombre et la

difficulté des questions posées aux ressources budgétaires et humaines disponibles pour l'évaluation. Ce dernier point constitue à lui seul une bonne raison d'investir dans la préparation du cahier des charges. L'ambition excessive de certains cahiers des charges est en effet une cause fréquente d'insatisfaction pour le commanditaire et de difficulté pour l'équipe d'évaluation. Il n'est pas rare de lire un cahier des charges qui demande une étude quantitative de multiples impacts et qui fixe en même temps un délai de quelques mois et une enveloppe budgétaire restreinte. A titre de comparaison, une étude portant sur une trentaine d'analyses d'impact réalisées en Irlande a montré que le coût moyen d'un traitement approfondi a été de l'ordre de 50 à 100 KEuro, et que la durée typique d'une telle analyse est de 6 à 12 mois.

Ouvrir l'évaluation à des partenaires plus nombreux

L'utilisation concrète des évaluations pourrait également progresser grâce à un meilleur équilibre entre l'indépendance de l'évaluation et son interaction avec les responsables du programme. L'expérience a en effet montré qu'une forte interaction entre l'équipe d'évaluation et les gestionnaires augmente la qualité et l'utilité de l'exercice.

Par ailleurs, il est fréquemment souhaitable d'élargir le groupe de pilotage à un plus grand nombre de partenaires pour accroître la légitimité et la pertinence des conclusions. Il convient cependant de bien maîtriser le nombre et le choix des personnes invitées à participer à l'évaluation, ainsi que l'efficacité du processus d'évaluation. L'organisation d'une véritable réflexion collective peut être organisée au sein d'un groupe de pilotage élargi en utilisant des modes de travail appropriés.

■ Impliquer les gestionnaires pour favoriser l'appropriation

L'appel à des équipes d'évaluation externes se justifie par la nécessité de recourir à des compétences dont l'administration ne dispose pas toujours, que ce soit à propos des techniques d'évaluation ou à propos d'un mécanisme d'impact particulier. L'externalité de l'évaluation a également l'avantage de permettre un regard nouveau, distancié, et impartial sur le programme évalué et de renforcer la crédibilité des résultats. C'est ce dernier avantage qui justifie l'exigence d'indépendance rappelée dans les différents textes de l'Union européenne relatifs à l'évaluation.

Le souci d'indépendance ne doit cependant pas aboutir à ce que l'équipe d'évaluation travaille dans l'isolement. Dans le domaine des Fonds structurels européens, les expériences convergent pour montrer que l'interaction entre l'équipe d'évaluation et les gestionnaires renforce la motivation de ces derniers, évite les problèmes d'accès à l'information et conduit à une meilleure utilisation des résultats. A contrario, la crédibilité de l'équipe d'évaluation peut être affaiblie si son indépendance s'accompagne d'une méconnaissance des systèmes administratifs, si elle suscite de la méfiance et donc des attitudes défensives.

Le manque de contact avec les gestionnaires étant toujours préjudiciable, il convient d'exiger une disponibilité suffisante de la part de l'équipe d'évaluation, et en particulier de la part des personnes qui se sont

nommément engagées à réaliser le travail. Un temps suffisant doit également être laissé pour l'interaction avec l'équipe d'évaluation à chaque étape intermédiaire.

Pour aider à dépasser les blocages, une règle de bon sens consiste à mettre en place des retours d'informations vers les gestionnaires pendant le déroulement de l'évaluation. Le groupe de pilotage peut faire circuler des notes intermédiaires et les premiers résultats des enquêtes ou analyses. La formule a l'avantage de permettre la correction d'erreurs factuelles qui sont aisément détectées par les gestionnaires. Dans au moins deux cas (Land de Brandebourg et Région Aquitaine), ces échanges d'information sont passés par des canaux informatiques.

Pour que l'exercice d'évaluation soit perçu comme plus impartial, il est également possible d'élargir le groupe de pilotage à plusieurs gestionnaires et/ou opérateurs du programme. Ces participants sont invités à parler en leur nom personnel, à respecter la confidentialité des points de vue exprimés en réunions et à ne pas se comporter en « représentant officiel » de leur institution. Cette pratique a des retombées positives en termes de qualité et de crédibilité des travaux, dans la mesure où les personnes invitées au groupe de pilotage peuvent faire entendre les attentes et les craintes qui s'expriment au niveau des gestionnaires. Elles peuvent aussi témoigner de l'impartialité des travaux d'évaluation.

Une autre recommandation consiste à concevoir l'évaluation, non pas comme la recherche systématique des problèmes et des échecs, mais comme un examen équilibré, attachant une importance suffisante à l'identification des succès et des bonnes pratiques, à leur explication et à leur mise en valeur. C'est de cette façon que l'on fera évoluer l'image de l'évaluation.

■ Ouvrir le groupe de pilotage à des partenaires extérieurs

● Instance d'évaluation pluraliste

La mise en œuvre des Fonds structurels européens est caractérisée par des cofinancements et des co-décisions entre plusieurs niveaux d'administration. Même si la responsabilité finale de la commande incombe presque toujours à un seul niveau, l'évaluation ne peut se passer des informations détenues aux autres niveaux, ni faire l'impasse sur les attentes

des autres niveaux. Dans ce contexte, il serait dangereux et inutile de considérer l'évaluation comme un exercice unilatéral commandé par un seul décideur pour son usage exclusif. Pour cette raison, la pratique conduit souvent à instituer un groupe de pilotage qui permet l'expression de plusieurs points de vue (voir page 69).

À partir du moment où il existe un groupe de pilotage et que celui-ci joue un rôle important dans l'évaluation, on peut se poser la question de son élargissement. On a vu plus haut qu'il est intéressant d'élargir le groupe à quelques gestionnaires et/ou opérateurs afin de rendre l'évaluation intéressante et crédible pour cette catégorie de partenaires. La même démarche peut-être adoptée pour procéder à d'autres élargissements à d'autres partenaires.

Plusieurs auteurs nord-américains recommandent d'élargir l'évaluation de façon à faire participer les principaux partenaires concernés. Les termes employés sont ceux de parties prenantes, protagonistes ou « porteurs d'enjeux » (stakeholders). On peut potentiellement associer à l'évaluation toute personne qui est potentiellement utilisatrice des recommandations, toute personne qui a un intérêt dans les informations produites, toute personne qui est susceptible d'être gagnante ou perdante dans le processus. On devine que les protagonistes d'une évaluation se comptent en grand nombre. Pour simplifier, il est possible de les répartir en grandes catégories dont les plus communes sont les financeurs, les gestionnaires, les opérateurs et les publics concernés.

L'élargissement du groupe de pilotage à une catégorie de protagonistes présente les mêmes avantages que ceux qui ont été cités plus haut à propos de la participation des gestionnaires : les personnes invitées font entendre les attentes et les craintes qui s'expriment à leur niveau, elles améliorent la pertinence des questions posées, et leur présence crédibilise l'évaluation. Une plus grande diversité au sein du groupe de pilotage met en présence un plus grand nombre de points de vue, facilite leur confrontation, et permet d'élever le degré de consensus. Chaque participant apporte ses questions, son système de valeurs, ses critères de jugement et ses normes. Sous réserve de savoir tirer parti de cette pluralité, l'évaluation aboutit à des conclusions plus riches, plus crédibles et plus utiles. La prise en compte des attentes de plusieurs partenaires permet une meilleure acceptation des résultats et augmente l'intérêt qui sera porté à ces résultats. Ces avantages sont synthétisés dans l'Encadré 48.

Encadré 48 - Les avantages d'une instance d'évaluation élargie

Constituer un Comité de pilotage de l'évaluation comprenant les différentes parties prenantes du programme permet d'assurer :

Une meilleure acceptation de l'évaluation par les évalués en créant des relations de confiance.

Un accès plus facile aux informations et une meilleure compréhension des faits et événements survenus au cours du programme sont nécessaires.

Des interprétations et recommandations qui tiennent compte de l'ensemble des points de vue importants.

Une diffusion des conclusions et une prise en compte des recommandations plus rapide et informelle.

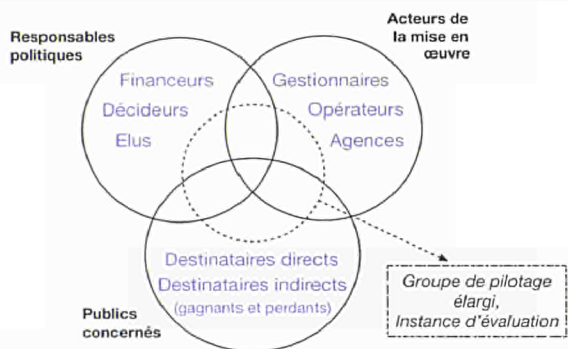
Les personnes à associer peuvent être recherchées parmi les universitaires, les chefs d'entreprise, les responsables d'organisations professionnelles, les leaders d'organisations non-gouvernementales, etc. Les participants sont toujours invités à titre personnel et promettent de respecter la confidentialité des points de vue exprimés en réunion. Dans le cas d'évaluations qui visaient à rendre des comptes au grand public, il a même été décidé d'associer des citoyens tirés au sort ou des journalistes. Si le groupe de pilotage est élargi à une dizaine de personnes porteuses de points de vue différents, son rôle dans le processus d'évaluation devient plus visible et plus déterminant. On le désigne alors par un autre terme qui reflète son importance, par exemple : comité d'évaluation, instance d'évaluation.

Pour composer une instance d'évaluation, il est important de lister les protagonistes susceptibles d'avoir un intérêt dans l'évaluation et dans ses résultats. On peut s'aider du schéma présenté dans l'encadré ci-dessous sous le nom de « sociogramme ». Ce schéma classe les protagonistes en trois catégories : les responsables politiques (élus, décideurs ou financeurs), les acteurs de la mise en œuvre (gestionnaires, opérateurs, agences d'exécution) et les publics concernés (destinataires directs ou indirects). On veillera à ce que les participants choisis soient disponibles (et assez motivés) pour participer régulièrement et activement aux différentes réunions du groupe et suivre les travaux.

La composition du Comité doit être la plus représentative possible des parties prenantes tout en demeurant de taille réduite (15 personnes maximum). Dans le cas des programmes européens on trouvera des représentants des Directions générales, ainsi que des instances nationales et locales de décision et de gestion concernées. Aux côtés de ces représentants institutionnels, on

fera participer des représentants des destinataires du programme. La question d'inclure ou non les gestionnaires et des représentants des opérateurs dans cette instance est souvent posée en termes de neutralité et d'indépendance de l'exercice d'évaluation. Il semble que les avantages de les inclure sont supérieurs aux risques d'une prise de pouvoir par ces personnes ; risques non négligeables dans la mesure où elles sont inévitablement plus motivées (c'est leur raison d'être) et plus compétentes techniquement que les autres membres. Un équilibre numérique entre les trois catégories d'acteurs offre des garanties à cet égard. On peut en outre prendre la précaution de confier la présidence à un responsable politique qui n'exerce pas une tutelle directe ou à une personnalité renommée (cette dernière solution est la plus fréquente dans les situations de partenariat multiple car elle évite toute suspicion de parti pris).

Encadré 49 – Sociogramme des personnes concernées par l'évaluation



Exemple : Le classement des protagonistes de l'évaluation d'un programme de développement rural

Certains acteurs peuvent se situer dans plusieurs des trois catégories précédemment définies. Les élus régionaux par exemple ont un rôle de décideurs, ils sont également porte-parole des publics concernés, et eux-mêmes bénéficiaires de financements. On trouve également des acteurs qui sont à la fois opérateurs et représentants du groupe destinataire, comme les Chambres d'Agriculture, qui constituent l'interlocuteur des pouvoirs publics et ont des liens forts avec les syndicats agricoles. Les administrations locales, pour leur part, font partie des trois catégories décideurs, opérateurs et publics concernés.

En relation avec ce qui est présenté ci-dessus, il est intéressant de s'arrêter sur la notion de partenariat, terme communément utilisé dans le vocabulaire européen depuis une dizaine d'années. Très proche de l'idée de partie prenante, il possède, dans le contexte de l'administration des Fonds structurels, un sens relativement spécifique en désignant des structures (partenaires) qui peuvent intervenir légitimement dans la définition et l'application des programmes financés par les Fonds structurels. La distinction faite plus haut entre les partenaires verticaux et les partenaires horizontaux est importante dans la mesure où seuls les premiers sont légitimes pour définir le mandat d'évaluation. Néanmoins, tous sont « parties prenantes » au programme et sont susceptibles d'être associés à la démarche d'évaluation, selon les modalités qu'il convient de préciser.

● Organisation de la réflexion collective

Lorsque le dispositif d'évaluation comprend une instance élargie, son travail est le même que celui d'un groupe de pilotage, à savoir : vérifier que des questions essentielles n'ont pas été oubliées, sélectionner les questions prioritaires, finaliser la rédaction des questions, préciser les critères et les normes de jugement, valider le cahier des charges, réagir sur les rapports intermédiaires de l'équipe d'évaluation, commenter les conclusions, valider les recommandations de l'équipe d'évaluation et émettre ses propres recommandations.

La différence entre un groupe de pilotage et une instance d'évaluation ne réside donc pas vraiment dans la nature du travail effectué, mais plutôt dans la façon d'effectuer ce travail. À l'inverse d'un groupe de pilotage, composé de quelques responsables administratifs qui ont l'habitude de travailler ensemble, une instance d'évaluation plus ouverte et plus nombreuse doit avoir des réunions très structurées pour produire un travail efficace.

Le dispositif généralement choisi est le suivant : le commanditaire désigne le président de l'instance dans le cadre de son mandat, ainsi qu'un médiateur. Sous la responsabilité du président qui dirige les débats, le médiateur assure de façon impartiale, le travail de préparation, d'animation et de compte-rendu des réunions. Le président et le médiateur doivent avoir un profil suffisamment neutre pour être acceptés par toutes les parties prenantes. Le médiateur doit avoir en outre une bonne connaissance technique de l'évaluation. Il peut appartenir à une cellule d'évaluation interne ou être recruté parmi les experts externes sur la base de quelques jours de travail.

Le médiateur prépare les réunions de travail, au besoin en réalisant des entretiens individuels avec tout ou partie des membres de l'instance. Il structure la discussion de façon à la rendre productive et conclusive. Il constate les éléments de consensus dans des comptes rendus. Le travail d'une instance d'évaluation pluraliste comporte des phases d'exploration au cours desquelles on privilégie l'expression des différences d'opinion (à propos des questions évaluatives, à propos des critères de jugement, à propos des recommandations, etc.). Ces phases exploratoires ont une dimension maïeutique, car le médiateur s'efforce de faire naître une expression de tous les points de vue, sans censurer les opinions minoritaires.

Les phases de créativité sont suivies par des phases de recherche de consensus qui conduisent, dans un esprit d'impartialité, à la production du travail collectif (finalisation du cahier des charges, validation des conclusions, production de recommandations). On trouvera dans le Volume 3 de cette collection, consacré aux outils d'évaluation, deux exemples de techniques de travail en groupe (METAPLAN®, vote coloré) qui permettent de conduire ces deux phases de travail de façon rapide, impartiale et consensuelle.

Faut-il qu'une évaluation soit consensuelle dans sa totalité ? Cette question devient importante lorsque le processus est élargi à de nombreux partenaires dont les points de vue sont diversifiés. La plupart des évaluations cherchent à produire une liste de conclusions acceptées par tous les partenaires, mais cela n'occulte-t-il pas le fait que la pluralité des points de vue et des opinions est un principe démocratique de base ?

Dans plusieurs expériences récentes, le processus d'évaluation a été construit de façon à accepter, et même à expliciter, les divergences de jugement entre partenaires. Une de ces expériences, conduite dans le cadre du programme MEANS, est décrite dans le Volume 4 de cette collection, qui présente plusieurs innovations techniques. Cette expérience repose sur le principe de l'analyse multicritère-multijuge. Comme son nom l'indique, cette technique permet de prendre en compte plusieurs critères d'évaluation et reconnaît que les membres d'une instance d'évaluation n'accordent pas les mêmes priorités aux mêmes critères. Par exemple, certains partenaires jugent mieux une mesure qui a un bon impact à court terme sur l'emploi alors que d'autres privilégient l'accroissement de compétitivité à long terme, les deux critères n'étant d'ailleurs pas forcément opposés. Au moyen d'entretiens individuels confidentiels avec plusieurs partenaires, l'équipe d'évaluation explicite leurs arbitrages entre les critères

d'évaluation. Malgré d'importants désaccords sur les poids à donner aux différents critères, les partenaires interrogés ont classé les actions du programme d'une façon en partie concordante.

En prenant le risque d'accepter des conclusions divergentes, l'évaluation a donc permis de révéler une zone de consensus, partiel mais véritable, sur certaines actions efficaces ou inefficaces. Cette issue est peut être préférable à celle d'une évaluation qui recherche une conclusion complètement consensuelle, mais qui risque d'apparaître comme factice.

Consolider les techniques d'évaluation

A l'heure actuelle la panoplie des outils classiques de l'évaluation reste mal adaptée à l'évaluation des programmes socio-économiques. Les raisons de ces difficultés techniques tiennent aux spécificités de ces programmes : imbrication des objectifs, multiplicité des interventions, dimension macroéconomique des principaux impacts attendus, attentes de plusieurs co-décideurs vis à vis de l'évaluation.

Cette collection récapitule et présente une série de progrès techniques applicables à court et à moyen terme et qui devraient permettre de mieux adapter les méthodes d'évaluation au contexte de la politique européenne de cohésion. Les principaux progrès réalisables passent par une collecte de données plus fiables, un meilleur niveau d'analyse des impacts et une plus grande capacité des équipes d'évaluation à déboucher sur des conclusions.

■ Gérer les cas d'objectifs ambigus ou imbriqués

Les programmes socio-économiques ont généralement des objectifs multiples. Au stade de l'évaluation ex ante, mais quelquefois aussi aux stades suivants, l'équipe d'évaluation est amenée à constater des ambiguïtés ou des incohérences dans ces multiples objectifs. Faut-il pour autant prendre ce prétexte pour déclarer le programme inévaluable et arrêter le travail à ce stade sur une conclusion critique ? Cette solution de simplicité est trop souvent adoptée par les équipes d'évaluation avec la permission des commanditaires.

Pour ne pas s'arrêter à cette limite, il faut que l'équipe d'évaluation devienne capable de juger les réussites et les échecs à l'aide de critères d'évaluation dépassant les objectifs, lorsque ceux-ci ne sont pas définis assez clairement. Une bonne solution consiste à poser aux gestionnaires et aux opérateurs la question suivante : « quels impacts (positifs / négatifs) pensez-vous que votre action va produire ? ». En réponse à cette question, on obtient une liste des "impacts attendus qui permettent de reconstruire la logique " du programme avec de meilleures chances qu'il soit évaluable. Cette démarche a souvent été employée avec succès en dehors du domaine des Fonds structurels. Des travaux pilotes, réalisées dans le cadre du programme MEANS, ont permis de la tester dans le cadre de l'évaluation

d'un programme de développement régional. Cette expérience est présentée dans le Volume 4 de la Collection. Elle emploie la technique de la cartographie conceptuelle qui permet d'associer efficacement plusieurs dizaines de gestionnaires et d'opérateurs au travail d'explicitation de la logique du programme.

L'expérience a porté sur le programme de développement économique cofinancé et co-décidé par la Commission européenne, le gouvernement espagnol et la région de Valencia. L'évaluation a débuté par une série de réunions au cours desquelles les partenaires ont identifié les impacts attendus et les ont rassemblés en groupes (ex : « diversification et efficience des ressources énergétiques »). Chaque groupe d'impacts a pu ainsi être défini de façon suffisamment précise pour construire des critères d'efficacité et pour choisir des indicateurs d'impact (ex : « nombre d'entreprises consommant du gaz naturel »).

Le caractère multisectoriel et multi-objectifs des programmes pose un autre problème d'évaluation qui est celui de la synergie. Les mesures ou les actions d'un programme sont généralement évaluées comme si elles étaient indépendantes les unes des autres, sans prendre en compte le fait que les impacts peuvent se renforcer mutuellement (effet de synergie), ou se contrarier (effet de synergie négative ou « d'antisnergie »). La synergie est rarement évaluée au niveau des résultats et des impacts pour les destinataires. Tout au plus évalue-t-on des facteurs de synergie comme la coordination administrative, la combinaison ou la superposition d'interventions.

Une proposition technique a été faite dans le cadre du programme MEANS. Elle comporte la construction d'une matrice des impacts croisés, la recherche des plus fortes synergies potentielles et la vérification de ces hypothèses de synergie par enquête auprès des destinataires. Cette démarche est présentée dans le Volume 4 de cette collection. Elle est conçue de manière à pouvoir être conduite sans accroissement significatif du volume du travail consacré à l'évaluation.

■ Mieux analyser les impacts

● Renforcer la fiabilité et la pertinence des données

De nombreuses équipes d'évaluation se sont plaintes de l'insuffisance des données de suivi, tant en quantité qu'en qualité. Des progrès sont incontestablement réalisables en termes de consolidation des systèmes de

suivi, et notamment en termes de choix et de quantification des indicateurs. Dans un exercice réalisé ex ante ou à mi-parcours, l'équipe d'évaluation joue d'ailleurs un rôle important en appréciant la qualité des indicateurs de programme et en faisant, si nécessaire, des recommandations pour les améliorer. Le Volume 2 de la Collection entre en détail dans les problèmes de quantification et propose une méthode d'appréciation de la qualité des indicateurs.

Des efforts considérables ont été faits à l'échelle européenne dans le sens d'une régionalisation des statistiques, dans la perspective de faciliter la quantification des indicateurs de contexte. Compte tenu de l'ampleur de la tâche et des investissements nécessaires, il sera probablement justifié de poursuivre ces efforts pendant les années à venir. Des perspectives particulièrement prometteuses s'ouvrent avec la généralisation des systèmes d'information géographique.

Des progrès sont également réalisables dans le sens d'une meilleure compréhension des rôles que peuvent jouer le suivi et l'évaluation dans la collecte des données. Les données de suivi sont collectées de façon nominative, systématique, exhaustive et permanente. Il est attendu que ces données soient complètement fiables en ce qui concerne les ressources mobilisées et les réalisations du programme. Il est plus difficile de collecter de telles informations à propos des résultats, et surtout à propos des impacts. Pour éviter trop de lourdeur, pour obtenir des informations plus fiables et pour ne pas donner une impression de bureaucratie, on préfère souvent collecter les données d'impact au moyen d'enquêtes ponctuelles, anonymes et limitées à de petits échantillons. Cette technique de collecte de données est celle que privilégie l'évaluation lorsqu'elle réalise une analyse approfondie. Il y a donc un juste équilibre à trouver entre l'amélioration des systèmes de suivi, d'une part, et le développement des évaluations approfondies, d'autre part.

● Améliorer l'analyse des impacts macroéconomiques

L'analyse des mécanismes d'impact macroéconomiques peut également être améliorée. Ce type d'analyse est une nécessité pour l'évaluation de grands programmes dont les principaux objectifs s'expriment à l'échelle d'une région ou d'un pays. C'est aussi une difficulté majeure car, malgré une décennie d'efforts, les techniques d'évaluation dans ce domaine n'ont pas encore atteint un niveau de maîtrise suffisant. En particulier, les recherches et les expériences devraient se poursuivre pour schématiser et

modéliser les effets d'offre (voir page 155) qui jouent un rôle essentiel dans la production des impacts durables en termes de développement. Ces effets sont encore mal pris en compte donc difficilement évaluables, mais font l'objet de travaux prometteurs dont on doit espérer qu'ils déboucheront à moyen terme sur des techniques généralisables.

● **Développer les évaluations thématiques**

L'évaluation tend à privilégier l'analyse des impacts qui correspondent aux objectifs des axes et des mesures du programme. Or l'évaluation ne doit pas négliger les attentes des autorités politiques aux niveaux supérieurs. En ce qui concerne l'échelon européen, il convient de prendre en compte un certain nombre de priorités communes que sont par exemple l'environnement, l'emploi, l'égalité des chances. Il est également attendu à l'échelon européen que les succès obtenus par certains programmes dans tel ou tel domaine d'intervention (ex : infrastructures de transport, formation professionnelle, R&D) fassent l'objet d'évaluations approfondies dans une perspective de reproductibilité.

Des améliorations sont possibles pour intégrer davantage ces priorités, encore trop souvent négligées dans les rapports d'évaluation. L'évaluation thématique peut être conçue soit au niveau d'un programme et établit alors une synthèse de l'ensemble des impacts possibles des différentes mesures du programme par rapport à un thème donné. Il est également envisageable d'analyser les impacts d'interventions semblables et conduire une évaluation thématique au niveau d'une région d'un Etat, voire de plusieurs Etats. Le Volume 5 de la Collection propose des solutions techniques pour la réalisation de ce type d'évaluation.

■ **Dégager des conclusions plus opérationnelles**

● **Comparer les performances**

Aujourd'hui, les équipes d'évaluation manquent de repères pour dire si un objectif est réaliste ou excessivement ambitieux (cas d'une évaluation ex ante) ou pour dire si un coût par emploi créé est raisonnable ou excessivement élevé (cas d'une évaluation ex post). Il en résulte des rapports d'évaluation qui ont du mal à conclure, ou dont les conclusions se concentrent sur des détails de mise en œuvre en éludant les questions de résultats et d'impacts.

Les progrès réalisables dans le domaine de la régionalisation des statistiques et de l'harmonisation des indicateurs donnent l'espoir de réduire, à terme, l'hétérogénéité des données utilisées pour l'évaluation des programmes. Cela ouvrira des perspectives intéressantes, quoique encore très nouvelles, pour multiplier les comparaisons entre programmes et pour construire des références de jugement. L'intérêt de telles perspectives mérite certainement que les efforts d'harmonisation des indicateurs soient poursuivis, notamment à propos des résultats et des impacts obtenus.

A plus court terme, il est également possible de progresser à travers des coopérations inter-régionales impliquant deux ou plusieurs régions européennes partenaires. Ces coopérations permettent de comparer les instruments d'intervention employés de part et d'autre, d'apprécier les actions selon des critères communs, de comparer l'efficacité respective des deux régions, et de terminer l'évaluation par des conclusions claires et acceptables. Cet exercice, qui s'apparente à la technique de l'étalonnage (benchmarking) a été expérimenté de façon intéressante par les régions Emilie-Romagne et Pays-de-Loire, dans le domaine du tourisme rural.

Les comparaisons pourraient également être plus fréquentes au sein de chaque programme. Elles permettraient à chaque évaluation de conclure plus clairement sur le fait que certaines mesures sont des succès et que d'autres ne le sont pas. Le Volume 4 de la Collection fait plusieurs propositions dans ce sens.

La première proposition consiste à utiliser plus fréquemment les techniques de notation. Ces techniques sont largement employées au Royaume-Uni pour la sélection des projets à financer par les Fonds structurels européens. La notation ne constitue donc pas une véritable innovation. Elle a d'ailleurs été employée avec succès dans le cadre de nombreuses évaluations, y compris pour des programmes co-financés par les Fonds structurels. La notation permet de formaliser, de façon comparable, le jugement porté sur plusieurs mesures au sein d'un même programme. A condition de reposer sur des critères explicites et d'être réalisée de façon impartiale, la notation aide l'équipe d'évaluation à produire de véritables conclusions.

La deuxième proposition consiste à employer l'analyse multicritère pour comparer entre elles les mesures ou les actions du programme évalué. Cette proposition a un caractère plus innovant car l'analyse multicritère est généralement employée pour évaluer des variantes d'un projet. L'emploi de

cette technique permet de réaliser un progrès dans le cas de l'évaluation d'ensemble d'un programme multi-objectifs. La démarche a été testée dans le cadre du programme MEANS (Voir le Volume 4 de la Collection) et son usage pourrait être développé.

● **Considérer l'utilisation de l'évaluation comme une priorité**

L'utilisation des évaluations est certainement le point qui peut faire l'objet des plus grandes améliorations. La probabilité d'utilisation étant liée à son appropriation par les utilisateurs, des progrès sensibles pourraient certainement être réalisés en impliquant mieux les gestionnaires dans le processus d'évaluation et en l'ouvrant à l'extérieur vers de nouveaux utilisateurs potentiels comme cela a été proposé plus haut (voir page 203).

Au-delà, on peut également envisager d'intégrer l'évaluation dans une perspective plus globale de gestion des performances des programmes. La recherche d'une meilleure performance est une préoccupation légitime, aussi bien au niveau de l'Union européenne qu'à celui d'un Etat membre ou au sein d'un programme. Ce point est présenté et développé dans le Volume 2 de la Collection.

Des progrès importants peuvent probablement être faits à court terme du point de vue de l'utilisation de l'évaluation comme instrument permettant de rendre des comptes aux institutions démocratiques et aux citoyens. A l'image des rapports de cohésion publiés régulièrement par la Commission européenne, chaque programme est susceptible d'utiliser les travaux d'évaluation pour rappeler quels problèmes ont été traités et pour montrer la contribution du programme à leur résolution. Cette culture démocratique de l'évaluation, actuellement limitée à l'Europe du Nord, mérite certainement de s'étendre pour mieux répondre aux attentes des citoyens et des contribuables, de plus en plus exigeants à l'égard des pouvoirs publics.

Annexes



Annexe 1

Processus d'évaluation ex ante, à mi-parcours et ex post

Les tableaux qui figurent dans cette annexe permettent d'organiser le processus d'une évaluation de A à Z en parcourant toutes les étapes décrites dans ce Volume. Trois situations d'évaluation sont présentées en parallèle : évaluation ex ante, évaluation à mi parcours, évaluation ex post.

Les considérations présentées dans cette annexe ont un caractère général et doivent être complétées en se référant aux notes et guides diffusés par les autorités nationales et européennes, lorsque de tels documents existent.

	Décider, concevoir et lancer l'évaluation		
	ex ante ◀	mi-parcours ◄	ex post ▶
Décider l'évaluation			
Définir le champ	Page 67	Page 67	Page 67
Préciser les motifs de l'évaluation et l'utilisation prévue	Page 68 Ex : améliorer la pertinence et la cohérence du programme	Page 68 Ex : proposer des réallocations budgétaires	Page 68 Ex : valider les meilleures pratiques
Prévoir la participation des principaux partenaires dans un groupe de pilotage	Page 69	Page 69 Inclure des questionnaires de mesures et des opérateurs (Page 203)	Page 68 Inclure des porte-parole des publics concernés (Page 204)
Formaliser la décision d'évaluer dans un mandat et/ou un programme pluriannuel d'évaluation	Page 198	Page 198	Page 198 Particulièrement utile pour ne pas « oublier » l'évaluation ex post
Rédiger le cahier des charges			
Faire exprimer les attentes des partenaires, sélectionner les questions évaluatives et les critères de jugement, préciser les questions à approfondir	Page 72 Principalement : raison d'être, pertinence et cohérence Eventuellement : efficacité potentielle, notamment macroéconomique	Page 72 Principalement : cohérence, efficacité et efficience des premiers résultats Eventuellement : pertinence compte tenu de l'évolution du contexte	Page 72 Principalement : efficacité et efficience des résultats et impacts Eventuellement : évolution des besoins des groupes concernés
Rappeler le cadre réglementaire et décrire le programme	Page 80 Proposition de programme	Page 80	Page 80 Y compris les modifications décidées pendant la mise en œuvre
Lister les connaissances disponibles	Page 81	Page 81 Y compris les données de suivi et l'évaluation ex ante	Page 81 Y compris les données de suivi et l'évaluation à mi-parcours
Vérifier l'évaluabilité des questions et définir les grandes lignes de la méthode	Pages 75 et 81	Pages 75 et 81	Pages 75 et 81
Fixer les règles du jeu, le calendrier et le budget indicatif	Pages 82 et 83	Pages 82 et 83	Pages 82 et 83

	Décider, concevoir et lancer l'évaluation (suite)		
	ex ante ◀	mi-parcours ⏮	ex post ▶
Lancer l'évaluation			
Fixer les compétences requises pour l'équipe d'évaluation, indiquer et la structure des offres à faire, faire valider le cahier des charges par le groupe de pilotage, lancer l'appel d'offre	Page 86	Page 86	Page 86
Sélectionner l'équipe d'évaluation	Page 87	Page 87	Page 87
Etablir le plan de travail de l'équipe d'évaluation avec le souci de la qualité finale du rapport	Page 181	Page 181	Page 181

	Réaliser l'évaluation		
	ex ante ◀	mi-parcours ⏮	ex post ▶
Examiner la logique du programme			
Analyser la stratégie et : apprécier sa pertinence, analyser les objectifs et apprécier leur clarté, examiner la cohérence interne et externe des objectifs	Page 93 L'appréciation de la pertinence, de la clarté des objectifs et de leur cohérence sont particulièrement nécessaires	Page 93 Examiner si la stratégie est restée pertinente au vu de l'évolution du contexte, montrer comment les gestionnaires et opérateurs comprennent les objectifs, faire les propositions de clarification si nécessaire	Page 93 Reconstituer les objectifs qui ont été poursuivis « de facto »
Examiner la cohérence entre les objectifs et les modalités de mise en œuvre : allocations budgétaires, modes d'interventions, critères de sélection des projets, répartition des responsabilités	Page 96	Page 96 Cet examen est particulièrement nécessaire	Page 96
Identifier les résultats et impacts attendus	Page 105	Page 105 Réaliser un diagramme logique des impacts (Page 99)	Page 105 Réaliser un diagramme logique des impacts (Page 99)
Examiner la qualité du système de suivi ainsi que celle du système d'indicateurs, vérifier que	Page 105 Cet examen se limite aux principes de suivi déjà connus. L'examen des indicateurs proposés est particulièrement nécessaire	Page 105 L'examen du système d'indicateurs existant est particulièrement nécessaire	Page 105

	Réaliser l'évaluation (suite)		
	ex ante ◀	mi-parcours ◀▶	ex post ▶
Examiner les effets du programme			
Sélectionner et exploiter les informations existantes	Page 124 Principalement : expertise, littérature, statistiques	Page 124 Y compris les données du suivi	Page 124 Y compris les données du suivi
Réaliser des enquêtes complémentaires	Page 129	Page 129 Pour approfondir la connaissance de certains résultats et de certains mécanismes d'impact (Page 161)	Page 129 Pour approfondir la connaissance de certains résultats et impacts (Page 161)
Réaliser des estimations de résultats et d'impact	Page 132 Principalement par extrapolation à partir des impacts d'interventions similaires (Page 135) Eventuellement : Estimation des impacts macro-économiques (Page 145)	Page 132 Y compris grâce à l'analyse plus approfondie de certains résultats et mécanismes d'impact (Pages 116 et 164)	Page 132 Y compris grâce à l'analyse plus approfondie de certains résultats et impacts (Pages 116 et 164) Eventuellement : Estimation des impacts macro-économiques (Page 145)
Formuler et valider les conclusions			
Interpréter les résultats des enquêtes et analyses avec l'aide du groupe de pilotage, préparer un jugement impartial sur la valeur du programme	Page 137 Ex : jugement sur l'ambition des objectifs et la probabilité de les atteindre	Page 137 Ex : jugement sur la contribution des différentes mesures à la réussite du programme	Page 137
Rédiger le rapport d'évaluation de façon claire et soignée en formulant de réelles conclusions et en répondant clairement aux questions évaluatives	Pages 140 et 214	Pages 140 et 214	Pages 140 et 214
Apprécier la qualité du rapport et le valider	Pages 142 et 181	Pages 142 et 181	Pages 142 et 181

	Communiquer et utiliser les résultats		
	ex ante ◀	mi-parcours ◀▶	ex post ▶
Communiquer les conclusions et les recommandations			
Discuter les propositions de recommandations formulées par l'équipe d'évaluation, et les améliorer	Page 177 Ex : recalibrage des objectifs, clarification des objectifs, amélioration du système d'indicateurs	Page 177 Ex : amélioration des modalités de mise en œuvre, amélioration du système de suivi, réallocations budgétaires	Page 177 Ex : mise en exergue des meilleures pratiques, généralisation des leçons apprises
Décider d'un plan de communication dès la validation du rapport	Page 175	Page 175	Page 175
Organiser un séminaire de présentation des conclusions et recommandations	Page 177 Ex : présentation aux partenaires impliqués dans la conception du futur programme	Page 177 Ex : présentation aux autorités responsables du programme	Page 177
Diffuser une publication et/ou une brochure, faire un communiqué de presse	Page 178	Page 178	Page 178 Ex : brochure de présentation des meilleures réussites
Suivre la prise en compte des conclusions			
Demander que l'autorité gestionnaire du programme se prononce explicitement sur les conclusions	Page 52 Intégrer les conclusions dans les documents de programmation, y compris les conclusions relatives à l'ambition des objectifs	Page 52 Intégrer les conclusions dans la gestion du programme, y compris les conclusions relatives aux réallocations budgétaires, au système d'indicateurs et à la sélection des projets	/
Intégrer les résultats dans des synthèses nationales ou européennes		Pages 52 et 57 Ex : rapport national sur la mise en œuvre des Fonds structurels	Pages 52 et 57 Ex : rapport de la Commission européenne sur la cohésion économique et sociale

Annexe 2

Guide pour la préparation d'un cahier des charges

Cette annexe a pour but de donner des indications méthodologiques pour l'élaboration d'un projet d'évaluation. Elles sont issues de l'expérience des évaluations menées dans les administrations et de notre collaboration avec le Conseil Scientifique de l'Évaluation (CSE, 1996). Elles doivent être adaptées à chaque situation considérée.

■ Origine de l'évaluation

Le cahier des charges précise dans quel cadre réglementaire l'évaluation prend place, qui a été l'initiateur du projet, et qui prend la responsabilité du « mandat politique » d'évaluation.

■ Délimitation et description du champ de l'évaluation

Le cahier des charges explicite le champ couvert par l'évaluation, et éventuellement les limites de ce champ. S'agit-il d'évaluer l'ensemble du programme ou seulement certains aspects ? S'agit-il d'évaluer les effets sur l'ensemble du contexte socio-économique ou sur un groupe social particulier ?

La description du champ couvre au minimum les thèmes suivants :

- Quels sont les objectifs du programme ? De quelle façon ont-ils été définis ? Quels sont les problèmes sociaux qui ont motivé, qui justifient, l'action publique ?
- Quels sont les effets recherchés ou attendus ? Quelles sont les groupes cibles ?
- Quels sont les ressources mises en oeuvre ? Selon quelles procédures ?

Si les objectifs sont mal définis ou devenus obsolètes, l'évaluation doit-elle reconstruire des objectifs implicites à partir des différentes logiques qui ont été à l'oeuvre. Dans ce cas, l'évaluation sera d'autant plus fructueuse qu'une synthèse préalable des objectifs, même limitée et partielle, sera présentée dans le cahier des charges.

Il convient à ce stade de tenter une « modélisation » de l'action publique évaluée, autrement dit de décrire la « théorie d'action » officielle. Quels

sont les enchaînements de cause à effets que les responsables de la politique évaluée avaient anticipés, fût-ce implicitement, pour préconiser les solutions retenues ? Exemple : telle incitation financière induit tel comportement de la part des entreprises, etc.

Le changement social que souhaite promouvoir les pouvoirs publics devra être décrit selon la grille en trois points : ressources mobilisées, réalisations opérationnelles, résultats et impacts produits. On s'efforcera de décomposer ce dernier point sous l'angle des objectifs immédiats de l'action d'une part et, d'autre part, des finalités c'est-à-dire de la justification ultime, l'idéal que l'on n'atteint jamais directement. Si nécessaire, le cahier des charges précise que l'équipe d'évaluation devra reconstruire la logique du programme tel qu'il a été mis en œuvre sur le terrain.

■ L'usage futur des résultats de l'évaluation

Définir les finalités de l'évaluation consiste à clarifier ce que l'initiateur attend de cette évaluation, pourquoi elle lui paraît opportune ou nécessaire. Un travail préalable de réflexion et d'interaction est indispensable car le gestionnaire direct de l'intervention évaluée ne dispose pas toujours du recul nécessaire pour orienter l'évaluation dans la bonne direction. La raison principale pour laquelle l'évaluation est demandée doit être explicitée. Veut-on infléchir la stratégie du programme ? Si oui, pourquoi ? Veut-on modifier les procédures de suivi ou de sélection des projets ? Veut-on réallouer des ressources ?

Dans le même esprit, il est souhaitable que le cahier des charges anticipe les décisions qui sont susceptibles d'être prises à l'issue de l'évaluation. Cette anticipation n'a pas pour objet de dicter les conclusions de l'évaluation ni de restreindre la capacité de l'équipe d'évaluation à remettre en cause les hypothèses posées dans le cahier des charges, mais plutôt de lui permettre de mieux connaître les préoccupations des commanditaires, pour éviter des malentendus quant à leurs attentes. Les échéances des décisions susceptibles d'être prises à l'issue de l'évaluation seront précisées lorsque ce sera possible.

Pour accroître autant que faire se peut l'utilité future de l'évaluation, il importe de comprendre le contexte dans lequel s'insère le programme. Le projet d'évaluation doit tenir compte des facteurs tels que : degré d'innovation de l'action, les conflits dont il est le siège, la familiarité des administrations avec la démarche d'évaluation ? Les intérêts et stratégies

des différentes parties prenantes doivent être recensés avec acuité pour comprendre les « ressorts » de l'action et les conditions à réunir pour que l'évaluation soit crédible et ses conclusions acceptées par les acteurs. Les éléments clés sont :

- L'action évaluée s'inscrit-elle dans la continuité ou en rupture vis à vis des politiques passées ?
- Quels sont les domaines d'incertitude et ceux pour lesquels les parties prenantes disposent de références communes ?
- Qui sont les « décideurs légitimes » ? Quels sont les opérateurs qui mettent en oeuvre l'intervention ?
- Quels sont les publics qui pourraient être bénéficiaires ou lésés par l'intervention ?

Les futurs utilisateurs de l'évaluation doivent également être identifiés précisément. On veillera à préciser la nature des relations entre acteurs (coopération, concurrence, etc.) en distinguant les trois niveaux : responsables politiques, acteurs de la mise en oeuvre et publics concernés.

■ Questions évaluatives et critères de jugement

Le cahier des charges pose de façon claire et concise un petit nombre de questions auxquelles l'équipe d'évaluation devra répondre dans ses conclusions.

Dans la mesure du possible, chaque question évaluative intègre un critère de jugement comme dans les questions suivantes :

- Est-ce que l'action se justifie au regard des besoins ou problèmes sociaux à résoudre ? (pertinence) ;
- Y a-t-il des conséquences inattendues, positives ou négatives ? ;
- Pourrait-on produire plus d'effets bénéfiques en adoptant d'autres dispositions de mise en oeuvre ? (efficacité) ;
- Pourrait-on produire plus d'effets bénéfiques avec les mêmes ressources ou pourrait-on obtenir les mêmes résultats en utilisant moins de ressources ? (efficience) ;
- Peut-on trouver une meilleure complémentarité (cohérence) ou peut-on produire plus de synergie au sein du programme et par rapport à des interventions publiques connexes ?

■ Connaissances et données disponibles

Le cahier des charges mentionne quelles sont les sources d'informations existantes, si l'on connaît déjà des systèmes d'explication des impacts ou des références de comparaison. Il précise sous quelle forme se trouvent les connaissances disponibles (données brutes, informations structurées dans une base de données informatique, données déjà analysées). Il précise également qui a la responsabilité de réunir l'information (information mise à la disposition de l'équipe d'évaluation par tel ou tel acteur, information à réunir par l'équipe d'évaluation elle-même).

■ Méthodes

Sans fixer les détails de la méthode à employer, le cahier des charges fournit les informations indispensables pour que les équipes candidates puissent faire des propositions méthodologiques raisonnables. Il précise par exemple :

- Quel est le niveau souhaité pour l'approfondissement de l'analyse d'impact ?
- Quelle est l'importance accordée à la crédibilité des conclusions, à la possibilité de généraliser les résultats, à l'appropriation de ces résultats par les acteurs ?
- Si nécessaire, quels sites ont déjà été choisis pour réaliser des enquêtes ou études de cas et pour quelles raisons ?
- Quels sont les indicateurs les plus importants du point de vue du commanditaire ?

Il convient également de présenter les principales hypothèses d'interprétation utilisables dans l'analyse de causalité :

- Pourquoi telle intervention semble-t-elle être une réussite ou un échec ?
- Quels sont les facteurs d'environnement ou facteurs exogènes (politiques connexes, contextes locaux...) à prendre en compte pour interpréter les effets observables ?

■ Dispositif d'évaluation

Il s'agit d'explicitier les rôles respectifs du commanditaire, du groupe de pilotage et de l'équipe d'évaluation. Le cahier des charges précise également la composition du groupe de pilotage et ses modalités de fonctionnement (composition, présidence, animation, secrétariat).

Le cas échéant, le cahier des charges précise s'il est attendu que l'équipe d'évaluation fasse des propositions pour élargir le groupe de pilotage et joue un rôle d'animation des réunions de ce groupe élargi.

■ Calendrier

Le cahier des charges fixe les différentes modalités de restitution des travaux de l'équipe d'évaluation (note intermédiaire sur une question particulière, séminaire, rapport final), ainsi que la date limite de remise de ces documents.

■ Ressources allouées

Une enveloppe budgétaire prévisionnelle est établie par le commanditaire. Le cas échéant, le montant de cette enveloppe est communiqué à toutes les équipes candidates qui seront principalement jugées sur la qualité et la pertinence de leurs propositions méthodologiques. Dans le cas contraire, le cahier des charges comporte des propositions méthodologiques détaillées et la sélection accorde une plus grande importance au prix.

■ Qualifications requises

Le cahier des charges précise les connaissances et l'expérience qui doivent être réunies dans l'équipe d'évaluation. Le cas échéant, il hiérarchise les qualifications par ordre d'importance.

■ Structure de la proposition

Pour faciliter la sélection de l'équipe d'évaluation, le cahier des charges impose aux équipes candidates de présenter leurs propositions selon une structure type, par exemple :

- Éléments du champ et du contexte de l'évaluation que le candidat considère comme essentiels.
- Rappel des questions posées et interprétation éventuelle de ces questions.
- Méthode d'évaluation proposée.
- Calendrier de travail détaillé.

- Composition nominative de l'équipe proposée, modalités de répartition du travail et de coordination au sein de l'équipe, références individuelles des membres de l'équipe.
- Prix.
- Références techniques, juridiques et financières de l'organisation contractante.

Annexe 3

Références bibliographiques

- Boyle, R., Lemaire, D. (eds), *Building Evaluation Capacity*, New Brunswick Publishers, (à paraître).
- Commission européenne (1998), *Les Fonds structurels en 1997. Neuvième rapport annuel*, Luxembourg: Office des publications officielles des Communautés européennes, ainsi que les rapports précédents.
- Commission européenne (1996), *Premier rapport sur la cohésion économique et sociale*, Luxembourg: Office des Publications officielles des Communautés européennes.
- Commission européenne (1998), *Evaluating EU Expenditure Programmes. A guide for Ex post and Intermediate Evaluation*, Luxembourg: Office des Publications officielles des Communautés européennes.
- Conseil Scientifique de l'Evaluation (1996), *Petit guide de l'évaluation des politiques publiques*, Paris :CSE.
- Cook, T.D., Leviton, L.C. and Shadish, W.R. (1985) 'Program Evaluation' in : Lindzey G. and Aronson E. (Eds.), *Handbook of Social Psychology*, Vol. I. (3rd ed.), New York : Random House.
- European Commission (1995), *Common Guide for Monitoring and Interim Evaluation of CSFs, SPDs and operations*, European Commission Structural Funds.
- GAO (1991), *Designing Evaluations*, Program Evaluation and Methodology Division, Washington
- Guba, E.G., Lincoln, Y.S. (1989), *Fourth Generation Evaluation*, Newbury Park: Sage Publications.
- Monnier, E. (1992), *Evaluations de l'action des pouvoirs publics*, Paris : CPE-Economica.
- Owen, M.O., Rogers, P. (1998), *Program Evaluation: Forms and Approaches*, London: Sage Publications.
- Rieper, O., Toulemonde, J. (Eds) (1997), *Politics and Practices of Intergovernmental Evaluation*, New Brunswick: Transaction Publishers.

Rist, R.C., Segsworth, R.V., et al. (1990), *Program evaluation & the management of government : patterns and prospects across eight nations*, New Brunswick : Transaction publishers.

Rossi, P., Freeman, H.E. (1993, 5th ed.), *Evaluation : A systematic Approach*, Newbury Park, CA : Sage Publications.

Vedung, E.(1997), *Public Policy and Program Evaluation*, New Brunswick: Transaction Publishers.

Index

A

additionnalité 32
 analyse approfondie 213
 analyse coût-avantages 57, 139
 analyse coût-efficacité 138
 analyse de régression 135, 167
 analyse déductive 164, 167
 analyse des données 189
 analyse inductive 164
 analyse multicritère 9, 139, 215
 analyse structurelle-résiduelle 155
 analyse SWOT 96, 139
 apport instrumental 25
 appréciation 56
 apprentissage 25
 apprentissage en double boucle 25
 apprentissage en simple boucle 25
 arbre d'objectifs 98, 99, 158
 attractivité 153
 avantage compétitif 150
 avantage concurrentiel 150

B

base économique 153
 besoin 25, 72, 91, 94, 137
 bien-fondé 93
 boîte noire 160
 budget indicatif 83

C

cadre logique 158, 164
 cahier des charges 70, 79, 81, 200
 calendrier 82
 capital humain 151
 cartographie conceptuelle des impacts 9, 212
 champ 67, 199
 co-décision 27
 coefficient 123

coefficient d'extrapolation 135
 cohérence 75
 cohérence externe 100, 105
 cohérence interne 97, 105
 collecte de données 131
 comité de suivi 69
 comité d'évaluation 21, 206
 conclusions de l'évaluation 141, 173
 contrôle 63, 72
 critère de jugement 73
 critère de sélection 104
 cycle de programmation 32, 91

D

démocratique 20, 21, 26, 47
 destinataire 129, 131, 206
 développement endogène 154
 développement exogène 153
 diagramme logique des impacts 99, 159, 160
 diffusion de l'évaluation 173
 dispositif d'évaluation 69, 208
 donnée primaire 125, 129, 132, 188
 donnée secondaire 125, 132, 188
 durabilité 75

E

échelle de notation 9
 effet attendu 211
 effet brut 118
 effet d'aubaine 118, 120, 136
 effet de demande 156
 effet de déplacement 118
 effet de réseau 154
 effet de stigmatisation 165
 effet de substitution 119
 effet d'offre 156, 214
 effet fournisseur 145
 effet inattendu 164
 effet multiplicateur 145

effet structurant 147, 156
efficacité 73, 137
efficience 73, 137
employabilité 152
endoformative 26, 179
entretien de groupe 131, 163
entretien individuel 163
environnement productif 151
équité 75
étalonnage 138, 215
étape 31
étude de cas 134
étude d'évaluabilité 75, 87, 201
étude préparatoire 200
evaluanda 67
évaluateur 20
évaluation à mi-parcours 32, 37, 38,
52, 55, 73, 74, 85, 92, 97, 100,
102, 104, 105, 108, 110, 113, 133,
136, 157
évaluation démocratique 21
évaluation d'ensemble 124, 131
évaluation ex ante 32, 36, 37, 38, 52,
68, 71, 73, 92, 93, 94, 97, 102,
104, 105, 108, 109, 113, 135, 180
évaluation ex post 32, 53, 55, 71, 74,
92, 97, 102, 110, 133, 136, 157,
214
évaluation externe 203
évaluation managériale 20
évaluation pluraliste 21
évaluation récapitulative 53, 179
évaluation transversale 140
évaluation thématique 9, 59, 62, 157
externalité 108, 154
extrapolation 127, 135

F

facteur concurrent 127, 164
facteur exogène 164, 167
finalité 80

G

questionnaire du programme 129
grand projet 57, 61, 104
groupe de comparaison 117, 120,
135, 166
groupe de contrôle 117
groupe de pilotage 69, 70, 71, 205
groupes appariés 166

H

hiérarchie des objectifs 97

I

impact 25, 62, 72, 91, 106, 135, 186
impact attendu 107
impact intermédiaire 160
impact spécifique 107
indépendance 20, 44, 85, 203
indicateur 8, 40, 64, 138, 213
indicateur de contexte 109, 126, 215
indicateur de programme 109
instance d'évaluation 204
instrument de mesure 137

L

logique du programme 91, 158
logique reconstruite 211

M

macroéconomique 57, 132, 145
maïeutique 209
mainstreaming 9, 59
mandat 198
matrice des impacts croisés 9, 101,
137, 212
médiateur 139, 208

mesure 61
méta-évaluation 182
METAPLAN® 139, 209
méthode d'évaluation 81, 187
méthodologie 201
microéconomie 57, 132, 145
mise en œuvre 101
motifs 68

N

norme 137
notation 138, 215

O

objectif 91
objectivité 20
opérateur 129, 207
outil 84

P

partenariat 32, 42
partie prenante 205, 208
pertinence 73, 94, 137
plan de communication 176
politique 31
porteur de projet 129
principe de concentration 103
priorité communautaire 53, 59
priorités 185
programme 31, 186, 189
programme de développement socio-économique 6
programme pluriannuel d'évaluation 54, 55, 197
projet 31
promoteur de projet 129
prospectif 31
protagonistes 205

public 61
public concerné 22

Q

qualité de l'évaluation 181, 184
question évaluative 72, 75, 76, 200

R

raison d'être 24, 93
rapport d'évaluation 140, 142
rapporteur 7
réalisation 91, 106
recommandation 68, 141
rendre compte 17, 26, 53
reproductibilité 75
responsable 26
ressource 91, 102, 105
ressources humaines 151
résultat 25, 26, 63, 72, 91, 106
rétroaction 197
rétroaction directe 25
rétroaction indirecte 25
rétrospectif 31

S

situation contrefactuelle 116, 146, 166
sociogramme 206
sources statistiques 126
stade 31
stratégie 91, 94
subsidiarité 25
suivi 38, 63, 72, 188
synergie 56, 61, 75, 136, 212
système de suivi 109
système d'information géographique 133, 213

T

termes de référence 79
théorie d'action 159
théorie explicative 167, 177
transparence 26, 35, 37, 53
transversal 62

U

utilisateur 173
utilisation cognitive 177

utilisation conceptuelle 177
utilisation normative 179
utilité 73

V

valeur ajoutée communautaire 26
validité de construction 121
validité externe 61, 122, 123, 128
validité interne 121
viabilité 75

Remerciements

La préparation de ce Volume s'est étalée sur plusieurs années, de 1993 à 1998. Les travaux préparatoires ont comporté les étapes suivantes :

- Plusieurs ateliers ont réuni des experts, en majorité européens, sur les thèmes suivants « évaluation des effets structurants », « organisation des évaluations intermédiaires » et « appréciation de la qualité des évaluations ».
- Ces trois ateliers ont donné lieu à la publication de cahiers dont le contenu est en partie repris dans ce Volume.
- Près de trois cents fonctionnaires de la Commission européenne ont participé à des séminaires de formation. Le contenu de ces formations a progressivement évolué en fonction des attentes des participants. Ce Volume est construit sur la base du contenu de cette formation.
- Deux rapports ont été produits en forme d'aide mémoire pour la réalisation d'une évaluation ex ante et d'une évaluation ex post. L'annexe 1 de ce Volume en reprend l'essentiel.

La coordination des travaux de ce volume a été assurée par Philippe Goybet et Anastassios Bougas (D.G. XVI.G.2 : Coordination de l'évaluation) et a impliqué plusieurs Directions Générales de la Commission Européenne, notamment les D.G. II, V, VI, XIV et XVI et en particulier Miguel Benito Alonso, Andrea Mairate, David Sweet et René-Laurent Ballaguy (ENA).

Ce Volume a été rédigé par le Centre Européen d'Expertise en Evaluation, notamment avec la participation de Pascale Ambrois et Céline Fontaine.

Les auteurs souhaitent remercier toutes celles et tous ceux qui ont contribué, soit directement, soit indirectement, à la production de ce Volume et en particulier les experts dont les noms suivent : H. Aalbu (Nordlands Forskning), J. Bachtler (EPRC, Glasgow), A. Blalock (Admiralty Inlet Consulting, Seattle), R. Boyle (Institute of Public Administration, Dublin), J. Bradley (ESRI, Dublin), R. Brancatti (Ministerio del Bilancio, Rome), E. Calderon (Universidad Politecnica, Madrid), E. Chelimsky (American Evaluation Society), Hagens J. (TERP, Amersfoort), P. Nijkamp (Vrije Universiteit, Amsterdam), B. Perret (Conseil Scientifique de l'Evaluation, Paris), O. Rieper (AKF, Copenhagen), L. Senn (Universita Bocconi, Milan), E. Stern (Tavistock Institute, Londres), L.I. Ström (SIR, Ostersund).

Commission européenne

Évaluer les programmes socio-économiques
Conception et conduite d'une évaluation

Luxembourg: Office des publications officielles des Communautés européennes

1999 — 239 p. — 14,8 x 21 cm

ISBN 92-828-6627-0

BELGIQUE/BELGIÉ

Jean De Lannoy
Avenue du Roi 202/Koreningslaan 202
B-1100 Bruxelles/Brussel
Tel. (32-2) 538 03 49
Fax (32-2) 538 06 41
E-mail: jean.de.lannoy@inboard.be
URL: <http://www.jean-de-lannoy.be>

La librairie européenne/
De Europese Boekhandel
Rue de la Loi 244/Wetstraat 244
B-1040 Bruxelles/Brussel
Tel. (32-2) 295 26 39
Fax (32-2) 735 06 60
E-mail: mail@europ.be
URL: <http://www.europ.be>

Moniteur belge/Belgisch Staatsblad
Rue de Louvain 40-42/Louvainsvlei 40-42
B-1000 Bruxelles/Brussel
Tel. (32-2) 552 22 11
Fax (32-2) 511 01 64

DANMARK

J. H. Schütz Information A/S
Hertskovvej 10-12
DK-2620 Albertslund
Tel. (45) 43 83 23 00
Fax (45) 43 83 19 69
E-mail: schuetz@schuetz.dk
URL: <http://www.schuetz.dk>

DEUTSCHLAND

Bundesanzeiger Verlag GmbH
Vertriebsabteilung
Amsterdamer Straße 192
D-50735 Köln
Tel. (49-221) 97 66 80
Fax (49-221) 97 66 82 78
E-mail: vertrieb@bundesanzeiger.de
URL: <http://www.bundesanzeiger.de>

ΕΛΛΑΔΑ/GREECE

G. C. Eleftheroudakis SA
International Bookstore
Panepistimiotou 17
GR-10564 Athina
Tel. (30-1) 321 41 80/1/2/3/4/5
Fax (30-1) 323 98 21
E-mail: eleebooks@net.gr

ESPAÑA

Boletín Oficial del Estado
Tráfalgar, 27
E-28017 Madrid
Tel. (34) 915 38 21 11 (Libros),
915 38 17 15 (Suscrip.)
Fax (34) 915 38 21 11 (Libros),
915 38 17 14 (Suscrip.)
E-mail: clientes@com.boe.es
URL: <http://www.boe.es>

Mundi Prensa Libros, SA
Castelló, 37
E-28001 Madrid
Tel. (34) 914 36 37 00
Fax (34) 915 75 39 98
E-mail: libreria@mundiprensa.es
URL: <http://www.mundiprensa.com>

FRANCE

Journal officiel
Service des publications des CE
26, rue Desaix
F-75727 Paris Cedex 15
Tel. (33) 140 58 77 31
Fax (33) 140 58 77 00
URL: <http://www.journal-officiel.gouv.fr>

IRELAND

Government Supplies Agency
Publications Section
4-5 Harcourt Road
Dublin 2
Tel. (353-1) 661 31 11
Fax (353-1) 475 27 80

ITALIA

Licosa SpA
Via Duca di Calabria, 1/1
Casella postale 552
I-50125 Firenze
Tel. (39) 055 64 83 1
Fax (39) 055 64 12 57
E-mail: licosa@tiscali.it
URL: <http://www.tbcc.it/licosa>

LUXEMBOURG

Messagerie du livre SARL
5, rue Rarissen
L-2411 Luxembourg
Tel. (352) 40 10 20
Fax (352) 49 06 61
E-mail: mail@mdl.lu
URL: <http://www.mdl.lu>

NETHERLAND

Sdu Servicecenter Uitgeverij
Christoffel Plantijnstraat 2
Postbus 2014
2500 EA Den Haag
Tel. (31-70) 378 96 80
Fax (31-70) 378 97 83
E-mail: sdu@sdu.nl
URL: <http://www.sdu.nl>

ÖSTERREICH

Manz'sche Verlags- und
Universitätsbuchhandlung GmbH
Kohlmarkt 16
A-1014 Wien
Tel. (43-1) 53 18 11 00
Fax (43-1) 53 18 11 67
E-Mail: bestelln@manz.co.at
URL: <http://www.manz.at/index.htm>

PORTUGAL

Distribuidora de Livros Bertrand Ltd.ª
Grupo Bertrand, SA
Rua das Terras dos Vales, 4-A
Apartado 60037
P-2700 Amadora
Tel. (351-1) 495 90 50
Fax (351-1) 496 02 55

Imprensa Nacional-Casa de Moeda, EP
Rua Marquês Sá da Bandeira, 16-A
P-1050 Lisboa Coodex
Tel. (351-1) 351 03 99
Fax (351-1) 353 02 94
E-mail: ed.incm@mail.telepac.pt
URL: <http://www.incm.pt>

SUOMI/FINLAND

Akateeminen Kirjakauppa/
Akademiska Bokhandeln
Keskuskatu 1/Centralgatan 1
PL1616 Helsinki/Helsingfors
FIN-00101 Helsinki/Helsingfors
P/tin (358-9) 121 44 18
F/tax (358-9) 121 44 35
Sähköposti: akateemi@akateeminen.com
URL: <http://www.akateeminen.com>

SVERIGE

BTJ AB
Traktörvägen 11
S-221 82 Lund
Tfn (46-48) 18 00 00
Fax (46-48) 30 79 47
E-post: btjev-pub@btj.se
URL: <http://www.btj.se>

UNITED KINGDOM

The Stationery Office Ltd
International Sales Agency
51 Nine Elms Lane
London SW8 5OR
Tel. (44-171) 873 90 90
Fax (44-171) 873 84 63
E-mail: ipa.enquiries@theso.co.uk
URL: <http://www.the-stationery-office.co.uk>

ISLAND

Bokabud Larusar Böndal
Skólavörðung, 2
IS-101 Reykjavík
Tel. (354) 551 56 50
Fax (354) 552 55 60

NORGE

Sveita Norge AS
Ostenovsen 18
Boks 6512 Etterstad
N-0606 Oslo
Tel. (47-22) 97 45 00
Fax (47-22) 97 45 45

SCHWEIZ/SUISSE/SVIZZERA

Euro Info Center Schweiz
c/o CSEIC
Stämpelbachstraße 85
PF 492
CH-8003 Zürich
Tel. (41-1) 365 53 15
Fax (41-1) 365 54 11
E-mail: eics@osec.ch
URL: <http://www.osec.ch/eics>

BÁLGARÍA

Europress Euromedia Ltd
59, Blvd Vitosha
BG-1000 Sofia
Tel. (359-2) 980 37 66
Fax (359-2) 980 42 30
E-mail: Milena@mbbox.at.bg

ČESKÁ REPUBLIKA

ÚSIS
NIS-prodejna
Havelská 22
CZ-130 00 Praha 3
Tel. (420-2) 24 23 14 86
Fax (420-2) 24 23 11 14
E-mail: nikopas@seznam.cz
URL: <http://usscr.cz>

CYPRUS

Cyprus Chamber of Commerce and Industry
PO Box 1455
CY-1509 Nicosia
Tel. (357-2) 68 95 00
Fax (357-2) 68 10 44
E-mail: demetrp@cpci.org.cy

ESTI

Eesti Kaubandus-Tööstuskoode (Estonian Chamber of Commerce and Industry)
Toom-Kooli 17
EE-0001 Tallinn
Tel. (372) 648 02 44
Fax (372) 648 02 45
E-mail: einfo@koda.ee
URL: <http://www.koda.ee>

HRVATSKA

Medietrade Ltd
Povla Haza 1
HR-10000 Zagreb
Tel. (385-1) 481 94 11
Fax (385-1) 481 94 11

MAGYARORSZÁG
Euro Info Service
Europa Ház
Márgitsiget
PO Box 475
H-1396 Budapest 82
Tel. (36-1) 350 80 25
Fax (36-1) 350 90 32
E-mail: euroinfo@matav.hu
URL: <http://www.euroinfo.hu/index.htm>

MALTA

Miller Distributors Ltd
Malta International Airport
PO Box 25
Luqa LQA 05
Tel. (356) 66 44 88
Fax (356) 67 67 99
E-mail: gwmd@usa.net

POLSKA

Ars Polona
Kratowskie Przedmiescie 7
Skrytka pocztowa 1001
PL-00-950 Warszawa
Tel. (48-22) 826 12 01
Fax (48-22) 826 82 40
E-mail: ars_pol@benvy.hsn.com.pl

ROMÂNIA

Euromedia
Str. G-ral Berthelot Nr 41
RO-70749 Bucuresti
Tel. (40-1) 315 44 03
Fax (40-1) 314 22 86

ROSSIYA

CEEC
60-letiya Oktyabrya Av 9
117132 Moscow
Tel. (7-095) 135 52 27
Fax (7-095) 135 52 27

SLOVAKIA

Centrum VTI SR
Nám. Slobody, 19
SK-81223 Bratislava
Tel. (421-7) 54 41 83 64
Fax (421-7) 54 41 83 64
E-mail: info@centrum.vti.sk
URL: <http://www.vti.sk>

SLOVENIJA

Gospodarski Vestnik
Dunajska cesta 5
SLO-1000 Ljubljana
Tel. (386) 613 09 15 40
Fax (386) 613 09 16 45
E-mail: europ@vestnik.si
URL: <http://www.vestnik.si>

TÜRKİYE

Dünya Infotel AS
100, Ya Mahalleesi 34440
TR-80050 Beşiktaş/Istanbul
Tel. (90-212) 629 46 85
Fax (90-212) 629 46 27
E-mail: infotel@dunya-gazetesi.com.tr

AUSTRALIA

Hunter Publications
PO Box 404
3067 Abbotsford, Victoria
Tel. (61-3) 94 17 53 81
Fax (61-3) 94 19 71 64
E-mail: jpdavies@ozemail.com.au

CANADA

Les Éditions La Liberté Inc.
3020, chemin St-Jean-Foy
G1X 3V5 Saint-Joe, Québec
Tel. (1-418) 656 37 63
Fax (1-800) 567 54 49
E-mail: liberte@medion.qc.ca

Renouf Publishing Co. Ltd
5369 Chemin Canotek Road Unit 1
K1J 8J3 Ottawa, Ontario
Tel. (1-613) 745 26 65
Fax (1-613) 745 76 60
E-mail: order dept@renoufbooks.com
URL: <http://www.renoufbooks.com>

EGYPT

The Middle East Observer
41 Sherif Street
Gen. J. Shosaleh Marg
Tel. (20-2) 392 69 19
Fax (20-2) 393 97 32
E-mail: malouda@meobserver.com.eg
URL: <http://www.meobserver.com.eg>

INDIA

EBIC India
3rd Floor, Y. B. Chavan Centre
Gen. J. Shosaleh Marg
400 021 Mumbai
Tel. (91-22) 282 60 64
Fax (91-22) 285 45 84
E-mail: ebic@iasbm01.vsnl.net.in
URL: <http://www.ebicindia.com>

ISRAËL

ROY International
41, Mishmer Hayarden Street
PO Box 13055
61130 Tel Aviv
Tel. (972-3) 649 94 69
Fax (972-3) 648 60 39
E-mail: roy@nervision.net.il
URL: <http://www.royint.co.il>

Sub-agent for the Palestinian Authority:

Index Information Services

PO Box 19502
Jerusalem
Tel. (972-2) 627 16 34
Fax (972-2) 627 12 19

JAPAN

PSI-Japan
Asahi Sanbancho Plaza #206
7-1 Sanbancho, Chiyoda-ku
Tokyo 102
Tel. (81-3) 32 34 69 21
Fax (81-3) 32 34 69 15
E-mail: books@psi-japan.co.jp
URL: <http://www.psi-japan.com>

MALAYSIA

EBIC Malaysia
Level 7, Wisma Hong Leong
18 Jalan Perak
50450 Kuala Lumpur
Tel. (60-3) 262 62 96
Fax (60-3) 262 61 98
E-mail: ebic-kl@mol.net.my

MEXICO

Mundi Prensa Mexico, SA de CV
Rio Pánuco No 141
Colonia Cuauhtémoc
MX-06500 Mexico, DF
Tel. (52-5) 533 56 58
Fax (52-5) 514 67 99
E-mail: 101545.2361@compuserve.com

PHILIPPINES

EBIC Philippines
19th Floor, PS Bank Tower
Sen. Gil J. Puyat Ave. cor. Tindalo St.
Makati City
Metro Manila
Tel. (83-2) 759 66 80
Fax (83-2) 759 66 90
E-mail: ecpcppm@globe.com.ph
URL: <http://www.ecpcppm.com>

SRI LANKA

EBIC Sri Lanka
Trans Asia Hotel
115 Sri chittampalam
A. Gardiner Mawatha
Colombo 2
Tel. (94-1) 074 71 50 78
Fax (94-1) 44 87 79
E-mail: ebic-sl@tin.com

THAILAND

EBIC Thailand
29 Vanissa Building, 8th Floor
Sri Chidom
Ploenchit
10330 Bangkok
Tel. (66-2) 855 06 27
Fax (66-2) 855 06 28
E-mail: ebicthk@kac15.th.com
URL: <http://www.ebictk.com>

UNITED STATES OF AMERICA

Bernan Associates
4611-F Assembly Drive
Lanham MD20706
Tel. (1-800) 274 44 47 (toll free telephone)
Fax (1-800) 265 34 50 (toll free fax)
E-mail: query@bernan.com
URL: <http://www.bernan.com>

ANDERE LANDER/OTHER COUNTRIES/ AUTRES PAYS

Bitte wenden Sie sich an ein Büro Ihrer Wahl/ Please contact the sales office of your choical/ Veuillez vous adresser au bureau de vente de votre choix

Office for Official Publications of the European Communities
2, rue Mercier
L-2985 Luxembourg
Tel. (352) 29 29-42455
Fax (352) 29 29-42758
E-mail: info.info@opoc.ec.eu
URL: <http://eur-op.eu.int>

La collection MEANS : “Évaluer les programmes socio-économiques”

Cette collection en six volumes met à la disposition d'un public élargi les enseignements du programme MEANS, lancé par la Commission européenne en vue d'améliorer et de promouvoir les méthodes d'évaluation. Cette publication répond au besoin de plus en plus manifeste de disposer de références solides pour traiter de façon efficace les problèmes d'évaluation rencontrés dans la mise en œuvre de l'action publique en général et en particulier dans le domaine des Fonds Structurels.

De ce point de vue, cet ouvrage constitue un guide méthodologique original qui fournit des solutions à des problèmes techniques et organisationnels rencontrés par l'évaluation auxquels la littérature existante apporte peu de solutions.

La collection s'adresse non seulement aux évaluateurs professionnels pour qui elle constituera un vade-mecum précieux, mais aussi aux non-spécialistes et en particulier aux autorités de gestion et aux membres des comités de suivi des programmes structurels.

Chaque volume est conçu de manière à être lu indépendamment des autres, mais tous sont complémentaires et traitent conjointement la totalité des aspects de l'évaluation des programmes socio-économiques :

- I. Conception et conduite d'une évaluation
- II. Choix et utilisation des indicateurs pour le suivi et l'évaluation
- III. Principales techniques et outils d'évaluation
- IV. Solutions techniques pour évaluer dans un cadre de partenariat
- V. Evaluation transversale des impacts sur l'environnement, l'emploi et les autres priorités d'intervention
- VI. Glossaire de 300 concepts et termes techniques



OFFICE DES PUBLICATIONS OFFICIELLES
DES COMMUNAUTÉS EUROPÉENNES

L-2985 Luxembourg