

Méthodologies d'évaluation des résultats et de l'impact de programmes publics : les trois leçons de base.

Maurice Baslé, économiste, CREM-CNRS-Université
Rennes 1.
Cereq.

L'évaluation dans le cycle de programmation des PPP et PPPP

- Il existe trois types d'évaluation de PP et programme, chacun associé à un stade précis dans la vie d'un programme :
- – L'évaluation « ex-ante » a lieu au moment de la conception d'un programme. C'est un processus d'évaluation servant à préparer et à vérifier le bien-fondé d'une proposition d'un nouveau programme avant son lancement. Sa fonction est notamment de vérifier l'adéquation des objectifs par rapport aux besoins, enjeux ou problèmes à résoudre (la « pertinence »,). A la Commission Européenne et dans plusieurs administrations nationales en Europe, l'évaluation ex-ante est obligatoire avant de lancer un nouveau programme. Un exemple d'une évaluation ex-ante au niveau européen est celle qui a mené à la conception des nouveaux programmes 2007-2013 de la DG Education & Culture de la Commission Européenne.
- On parle d'évaluation stratégique ex ante quand l'exercice consiste à vérifier le respect de conditionnalités politiques (promesse d'effets d'égalité hommes/femmes par exemple) ou environnementales (moins d'empreinte carbone par exemple).
- –



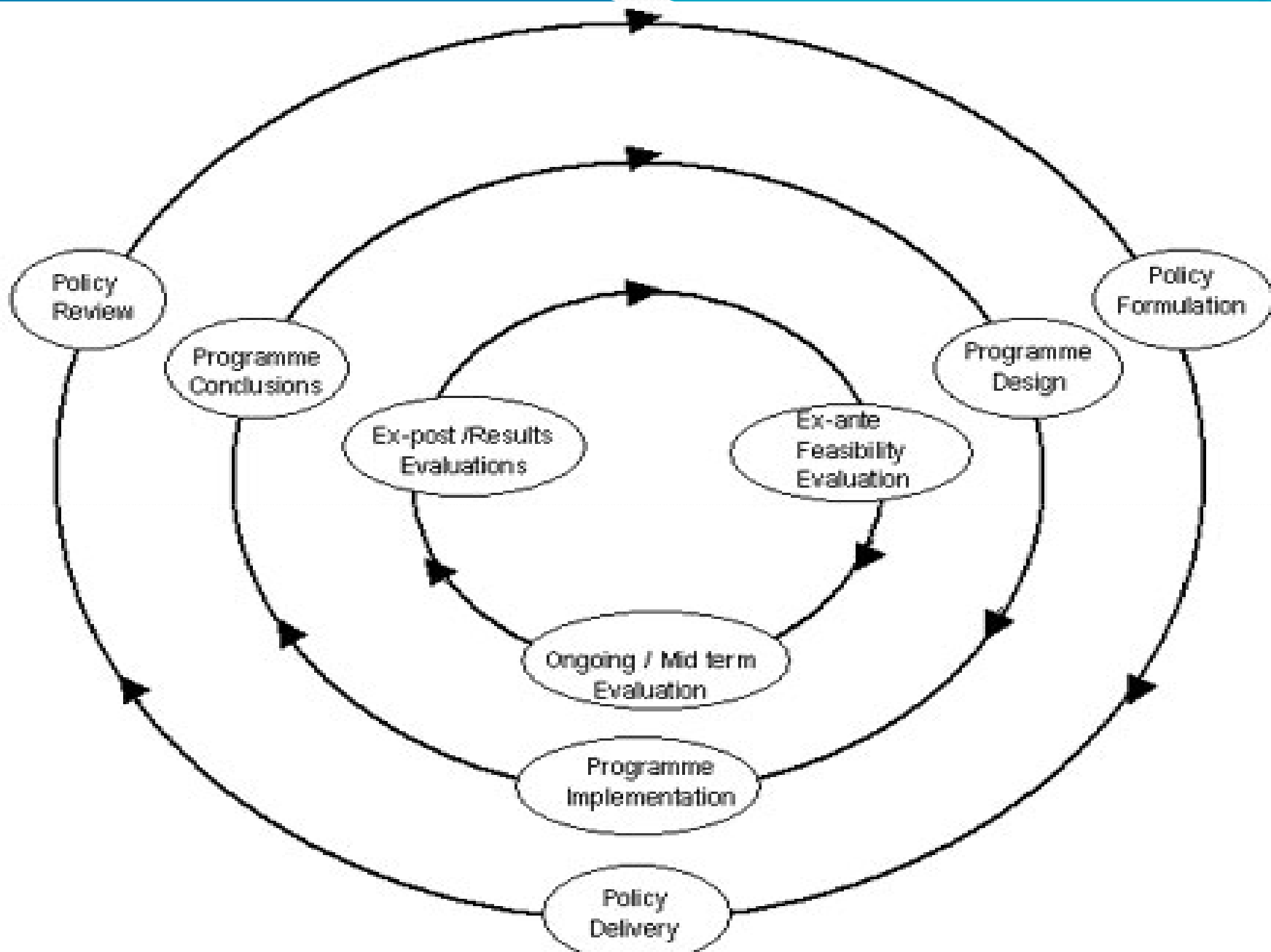
Méthodologies d'évaluation des résultats et de l'impact de programmes publics : les trois leçons de base.

- L'évaluation à mi-parcours (ou « intermédiaire ») peut être mise en œuvre pour vérifier, au milieu du cycle du programme, si les besoins sont toujours présents, si la gestion du programme se déroule comme prévu ou nécessite une amélioration, et pour analyser, éventuellement, les premiers effets du programme. Un exemple d'une évaluation à mi-parcours est celle, réalisée en 1999, du PREDIT₂, le Programme (interministériel) de recherche, de développement et d'innovation dans le domaine des transports terrestres. Cette évaluation avait permis plusieurs ajustements du programme, comme, par exemple, une meilleure coopération de ses différents « groupes thématiques », auparavant relativement cloisonnés.
- –L'évaluation chemin faisant qui s'apparente à l'ambition d'une amélioration en continu: on met en place du suivi (monitoring) sur les réalisations. On prévoit des cibles et on met en place des alertes...On parle aussi de pilotage stratégique. C'est la « tendance ». Elle va de plus en plus avec l'idée de PPP expérimentaux (donc sous « cloche » à tout moment).

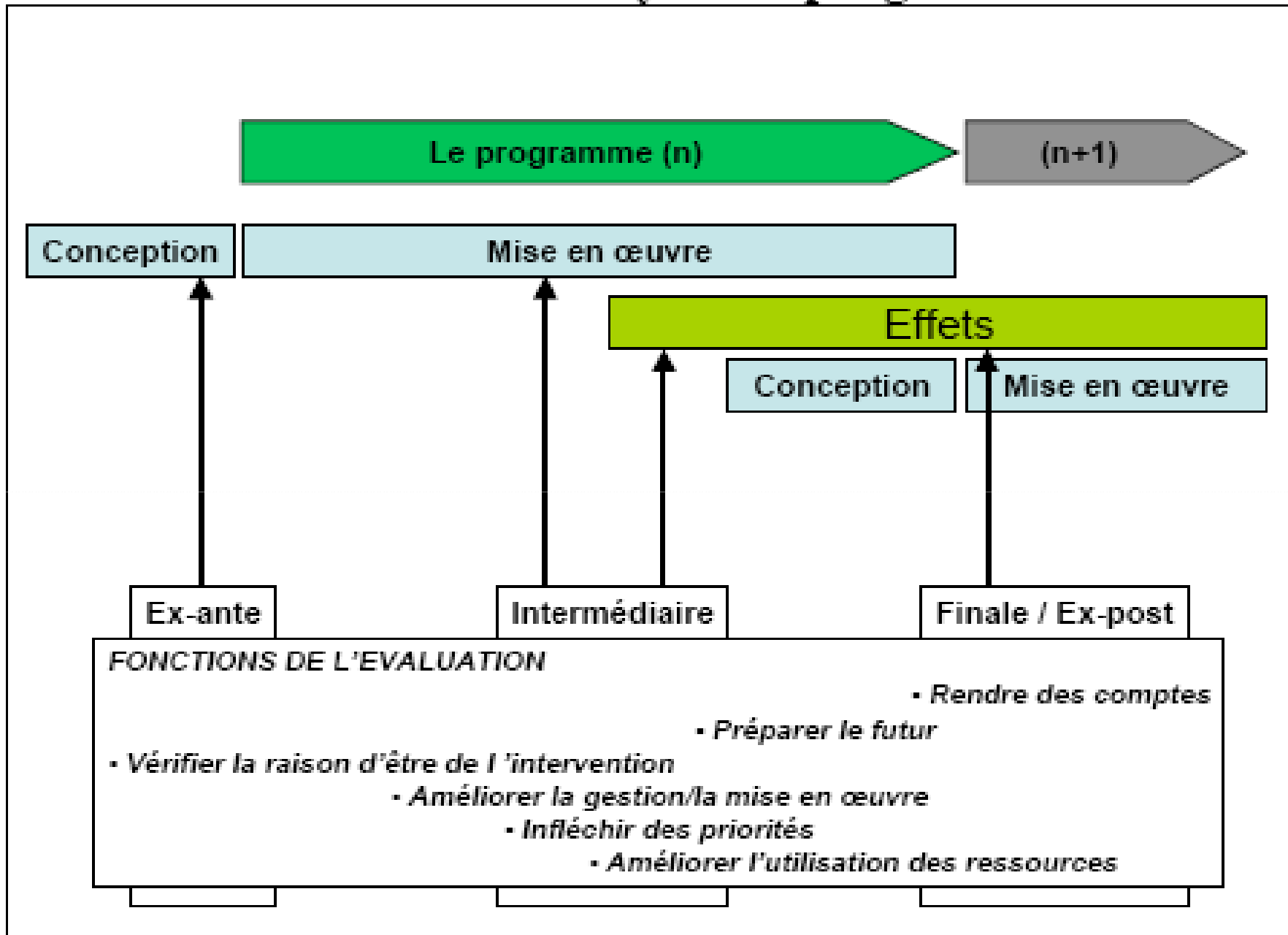


Méthodologies d'évaluation des résultats et de l'impact de programmes publics : les trois leçons de base.

- L'évaluation finale (en fin de programme) ou ex-post (après l'arrêt du programme) est, plus que les deux autres types d'évaluation qui sont davantage « managériales », réalisée le plus souvent dans le but, politique, de re-conception des PPP. Elle peut être accompagnée d'une communication de type « gouvernance participative ».
- Pour des raisons d'objectivité et de crédibilité, elle est dans ce cas le plus souvent réalisée en externe (panel d'experts, bureau d'études, universitaires). Dans l'idéal elle doit intervenir avant la phase de reprogrammation suivante.



L'évaluation dans le cycle de programmation





Trois leçons que je sais d'elles

- 1. Tous les PPP ou PPPP ne sont pas rendus facilement évaluable! Il y a donc un travail préliminaire et il serait utile de rendre les PPP plus stratégiques. Il faut quelquefois ré-écrire les objectifs (pas de « goal-free evaluation »). Un des items de l'évaluation est de s'interroger sur la pertinence des PPP (interrogation sur le « sens » et sur les objectifs par rapport aux questions socio-économiques)
- 2. Tous les indicateurs appelés au secours des résultats d'une évaluation ne sont pas également utilisables!
- 3. Il est préférable de recouper les approches et les résultats même avec les évaluations expérimentales!

Rappel des pré-requis du monitoring (suivi) et de l'évaluation des politiques publiques et des programmes cf

<http://ec.europa.eu/europeaid/evaluation/methodology/tools>

Il faut avoir le référentiel des PPP avant de se poser la question des indicateurs statistiques susceptibles d'être utilisables pour les évaluations de programmes.

Dans toutes les évaluations et dans le monitoring, le problème pudiquement passé sous silence est que dans les PPP, tout n'est pas d'emblée « raisonnablement » évaluable.

Dans les activités évaluatives nouvelles, il y a, il devrait y avoir des préalables.

1. Tous les PPP ou PPPP ne sont pas directement évaluable!

- « *L'absence d'objectifs bien définis est d'ailleurs, en même temps qu'un frein à la démarche d'évaluation, un constat fréquent des travaux d'évaluation. [...] L'énoncé des objectifs des politiques publiques dans les lois ou règlement se contente souvent, quand de tels objectifs sont mentionnés, de déclarations d'intentions très vagues. Elle se prolonge au niveau budgétaire par l'absence dans les documents produits à l'appui des demandes d'autorisation de crédits (les « bleus ») d'indications précises sur les objectifs des financements sollicités* » (Bourdin, André, & Plancade, 2004).
- Les PPP devraient toujours être construits comme des stratégies, avec des objectifs stratégiques et comme des déroulements d'activités utilisant des moyens humains et budgétaires au service des réalisations en rapport avec les objectifs.

Objectif

Volonté, ou Priorité

La notion d'objectif manipulée ici est celle d'une ligne directrice à suivre dans un référentiel de valeurs. Un verbe traduit le volontarisme qui est au service de la réalisation de l'objectif (renforcer l'égalité par exemple)

1. Tous les PPP ou PPPP ne sont pas directement évaluable!

- Arbre des Objectifs

L'Arbre des priorités (d'actions)

Structuration hiérarchisée d'objectifs, sous forme d'arborescence d'objectifs .

On distingue par convention des objectifs stratégiques, intermédiaires (par exemple développer l'apprentissage) et opérationnels (multiplier par deux les contrats d'apprentissage).

On peut, en vue du pilotage ou de l'amélioration continue, définir des cibles de résultats des programmes sur les objectifs (par exemple. Cible sur le nombre de contrats réalisés)

Méta-objectif stratégique

Les valeurs

Perception visionnaire pour la collectivité ou le gouvernement national, son rôle, son utilité sociétale « aider une jeunesse active »

Objectif stratégique

Les volontés

Orientation générale élaborée en réponse à un diagnostic, impulsée sur un axe fort d'engagement (« aider les jeunes en formation »)

Objectif intermédiaire

Les priorités

Grandes lignes identifiées comme prioritaires, guidant l'activité des responsables de programmes (« développer les contrats d'apprentissage »)

Objectif opérationnel

(L'agenda)

Déclinaison opérationnelle guidant les acteurs de programme dans l'exercice de leurs fonctions managériales

Il faut lier l'arbre des objectifs aux diagrammes logiques d'impact

« Le dialogue entre résultats et objectifs a cruellement manqué de formalisme ; le Diagramme Logique d'Impact et l'Arbre des Objectifs ont souvent été utilisés indépendamment, ce qui a mené de nombreux organismes à construire des « usines à gaz » générant une multitude d'indicateurs sans plan de circulation de l'information et sans stockage cohérent. » (Huauilmé, M.)

Le DLI ou Logigramme est l'outil du management basé sur les activités. Il permet de formaliser la manière par laquelle la volonté des acteurs se traduit sur le terrain. On pose une convention de causalité entre réalisations et premiers résultats, c'est-à-dire qu'on décrit la fabrication des résultats observés. Les réalisations, codées par objectif, produisent des effets sur les bénéficiaires directs ou indirects, et à plus ou moins long terme, on écrira la probabilité d'impact : les résultats probables les plus généraux et les impacts potentiels plus systémiques.



Avec les dli, on doit se mettre d'accord, au moins un moment, sur les conventions suivantes :

Comment la volonté se traduit-elle en réalisations ? Les intentions d'actions se traduisent-elles en réalisations effectives?

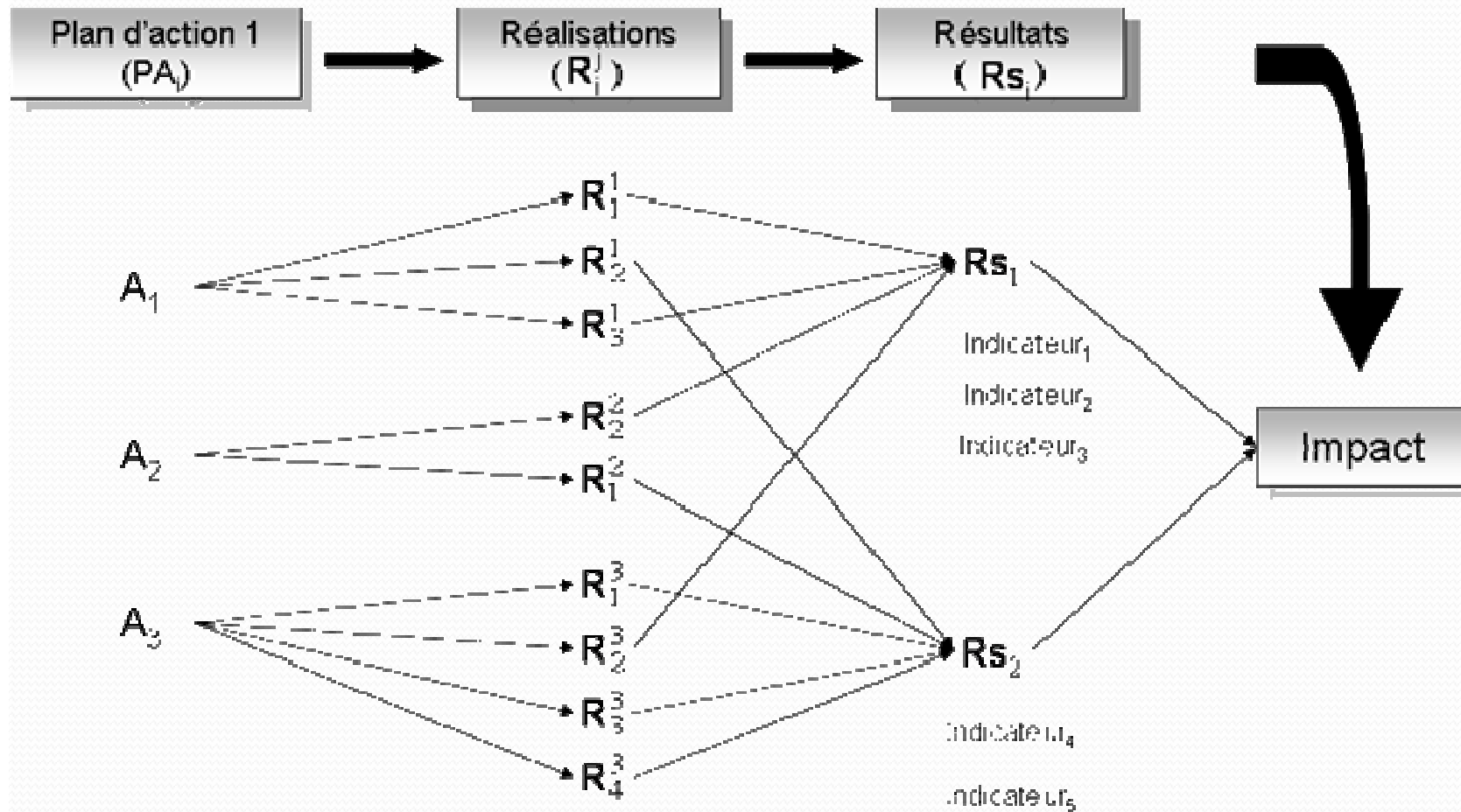
Peut-on et comment imputer les premiers résultats aux réalisations effectives ? Hypothèses de causalité.

Est-il probable que tel ou tel résultat ou ensemble de résultats finisse par avoir tel ou tel impact.

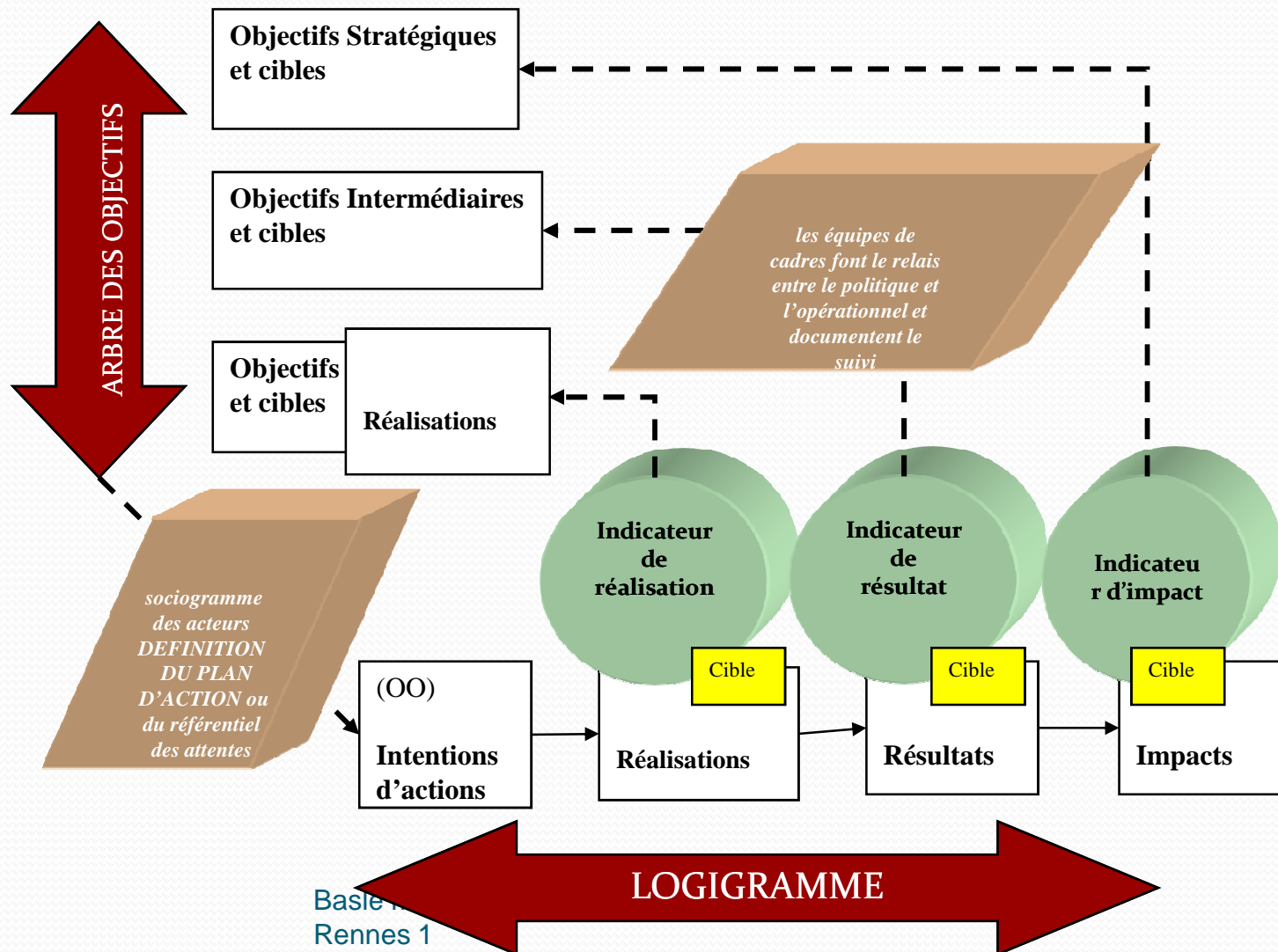
Spécification supplémentaire.

Ici on s'intéresse à des impacts potentiels sur des populations ou des territoires (ancrage territorial de populations, d'entreprises etc..)

La fabrique, le DLI ou logigramme



ARTICULATION *Arbre des objectifs et Logigramme*



temps les PPP n'ont pas été rendus évaluable!

- On les évalue quand même en laissant implicite tout ce qui a été dit précédemment.
- On va par exemple directement mesurer une évolution dans un taux d'insertion à l'issue d'une formation (tacitement ceci signifie qu'on a « caché » un logigramme et que le résultat est « rattaché à un objectif opérationnel qui lui-même est rattaché à un arbre des objectifs bien structuré au plan des priorités, et que cet arbre est cohérent avec l'arbre des valeurs que la politique est censée avoir comme référentiel »!

2. Tous les indicateurs ne sont pas des Indicateurs d'évaluation.

Beaucoup commencent les évaluations en commençant par définir les indicateurs qui vont servir de points de « notation » des résultats des PPP.

Il faut « finir » par les indicateurs et ne pas commencer par eux!

- Faute de « monitoring » –voir plus haut–au plus près des programmes, faute de données de suivi ad hoc, on va pointer vers des bases existantes.
- On confond souvent les données de contexte avec les données de suivi du programme, on remplace ces dernières, qu'on n'a pas, par les premières, qu'on a!

Les indicateurs de contexte s'appliquent à l'ensemble d'un territoire, d'une population ou d'une catégorie de population. L'indicateur de contexte n'informe pas sur l'efficacité du programme car il ne s'applique pas à la mise en œuvre du programme et à ses effets. Il s'applique toujours à l'ensemble du territoire éligible ou du public visé sans faire de distinction entre ceux qui ont été atteints par le programme et ceux qui ne l'ont pas été.

Les indicateurs de programme concernent seulement une partie ou la catégorie du public qui a été effectivement atteinte. Les indicateurs de programme cherchent à suivre à la trace les effets du programme, directs et indirects. Ils peuvent montrer que telle ou telle intervention est un succès ou que telle autre est un échec.



Les indicateurs de contexte sont construits sans but, sans référence à un PPP, sans idée de l'utilisateur, sans connaître les préférences de ceux-ci.

Au contraire, t"he purpose of programs indicators is to support effectiveness throughout the processes of planning, implementation, monitoring, reporting and evaluation – in other words, throughout the full spectrum of results-based Management.

Indicators may be used at any point along the results chain of inputs, activities outputs, outcomes and impacts.”

European Commission

Exemple.

Cas des indicateurs territoriaux européens ESPON, European spatial planning operational network.

La méthode d'évaluation des effets reste un peu « boîte noire » au plan pédagogique.

La première faiblesse fondamentale réside dans le fait que les indicateurs ne sont accrochés ni à des arbres des objectifs ni à des logigrammes. On est plus proche d'indicateurs d'évolution des contextes socio-économiques. Une deuxième faiblesse fondamentale est dans la question de la pédagogie des indicateurs : une annexe méthodologique définit des matrices d'interactions qualitatives. La méthode des effets suppose des tableaux régionaux d'input-output avec des conventions de causalité entre réalisations et premiers résultats. On ne décrit pas sur le mode probabiliste les liens entre résultats et impact synthétique?



Exemple de bonnes données de contexte :
exemple de
l'observatoire de la Diact

http://www.territoires.gouv.fr/indicateurs/portail_fr/index_fr.php

Quels sont les indicateurs clés des dynamiques et de la cohésion territoriales ?

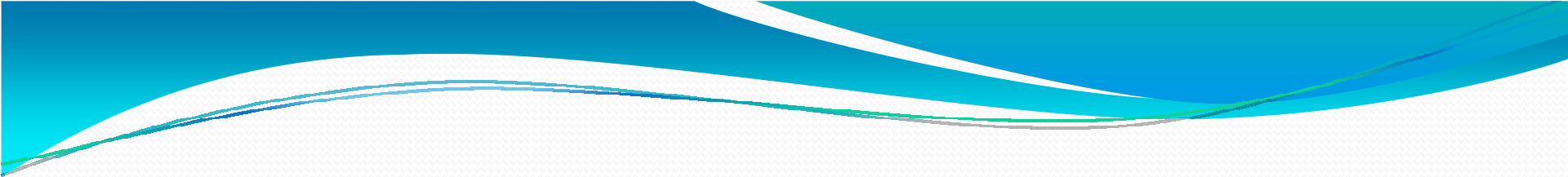
Comment se situent les régions françaises dans
Quels zonages et périmètres associés
aux politiques d'aménagement du territoire ?

A quels zonages appartient ma commune ?

Quelles spécificités des territoires littoraux ?


Quelles évolutions marquantes dans les domaines de la démographie, de
l'urbanisme, des activités économiques (agricoles, touristiques, industrielles...), de
la protection des espaces naturels ?

l'Union européenne ?



Beaucoup d' indices sont des statistiques, des données d'observatoires nationaux, régionaux ou même locaux mais pas des résultats d'évaluations menées au plus près des résultats des programmes.

« Un indicateur de contexte se définit comme une donnée qui fournit une base simple et fiable pour décrire une variable de contexte. Il informe sur la situation et son évolution dans le pays, dans la zone [...]. [...]. Les indicateurs sont produits par les services statistiques nationaux ou par certains programmes. » (European Commission, 2005)

- 
- Voici ce qui est nécessaire :quand on veut non seulement décrire le contexte mais évaluer les effets des PPP.
 - On ne considère pas que c'est l'ensemble des PPP et uniquement eux qui font évoluer le contexte!
 - On pose que les évolutions de contexte ont des causes multidimensionnelles parmi lesquelles, dans un sens ou dans un autre, il peut y avoir les PPP, mais aussi les comportements privés.
 - Plus on descend au plus près d'un Programme spécifique et indépendant, plus les résultats sont influencés par le PPP, plus la politique publique est complexe, plus le lien de causalité est ténu, plus les indicateurs d'observatoire sont inappropriés.
 - On considère les cibles d'objectifs, on considère les résultats et on met en rapport les résultats causés par le PPP avec la cible d'objectifs.



Les indicateurs de résultats se trouvent rarement dans les bases de données, les observatoires, surtout quand on descend au niveau régional et local.

On doit donc les créer spécifiquement pour chaque programme.
Et à chaque échelle territoriale et de la population-cible.

- **Conseils sur les indices:**
 - commencer par le moins ambitieux :
- Suivre les résultats directement imputables à des réalisations effectives, puis les familles de résultats,
 - établir ensuite alors des profils de résultats,
 - Suivre ces profils de résultats simples,
- ne construire une évolution d'indice synthétique d'impact qu'en dernier recours!

● Indicateur d'impact et d'utilité



On se situe dans la dimension des objectifs stratégiques, mais c'est un impact ou une utilité qui peut résulter d'un effet-système (avec beaucoup de causes entremêlées).
C'est l'indicateur le plus compliqué à produire, attention à ne pas prendre ses désirs pour des réalités ; on peut cependant se contenter d'approximations, des indicateurs existants qu'il faudrait décontextualiser.

Exemple 1 : l'effet-emploi d'un ensemble de programmes locaux dans un quartier sur plusieurs années : il y a plusieurs causes, des synergies entre les causes, des effets qui se compensent et donc des programmes sans effet net global

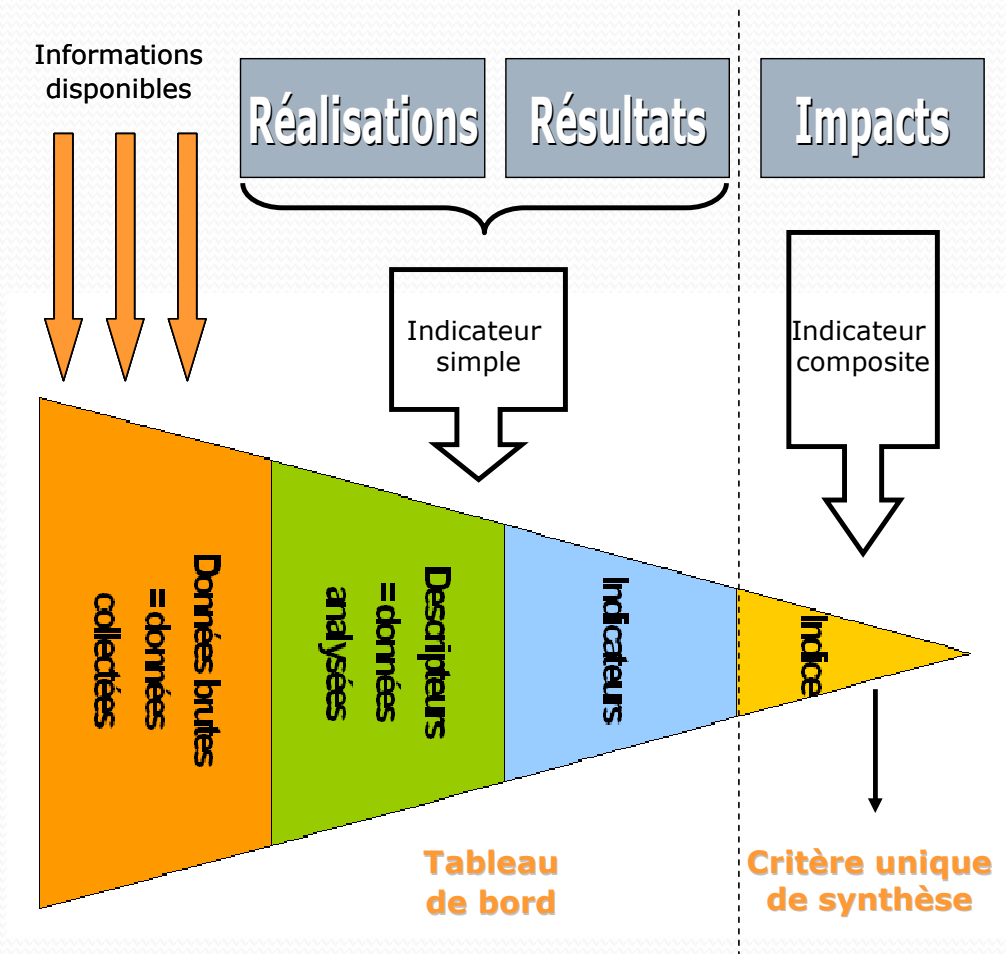
Exemple 2 : l'utilité sociale a besoin d'être précisée au niveau fin, utilité pour qui, utilité en quel sens, dans quelle dimension exacte (bien être, réduction des inégalités, acquisitions de droits, développement des domaines de libertés d'action)

Quitter

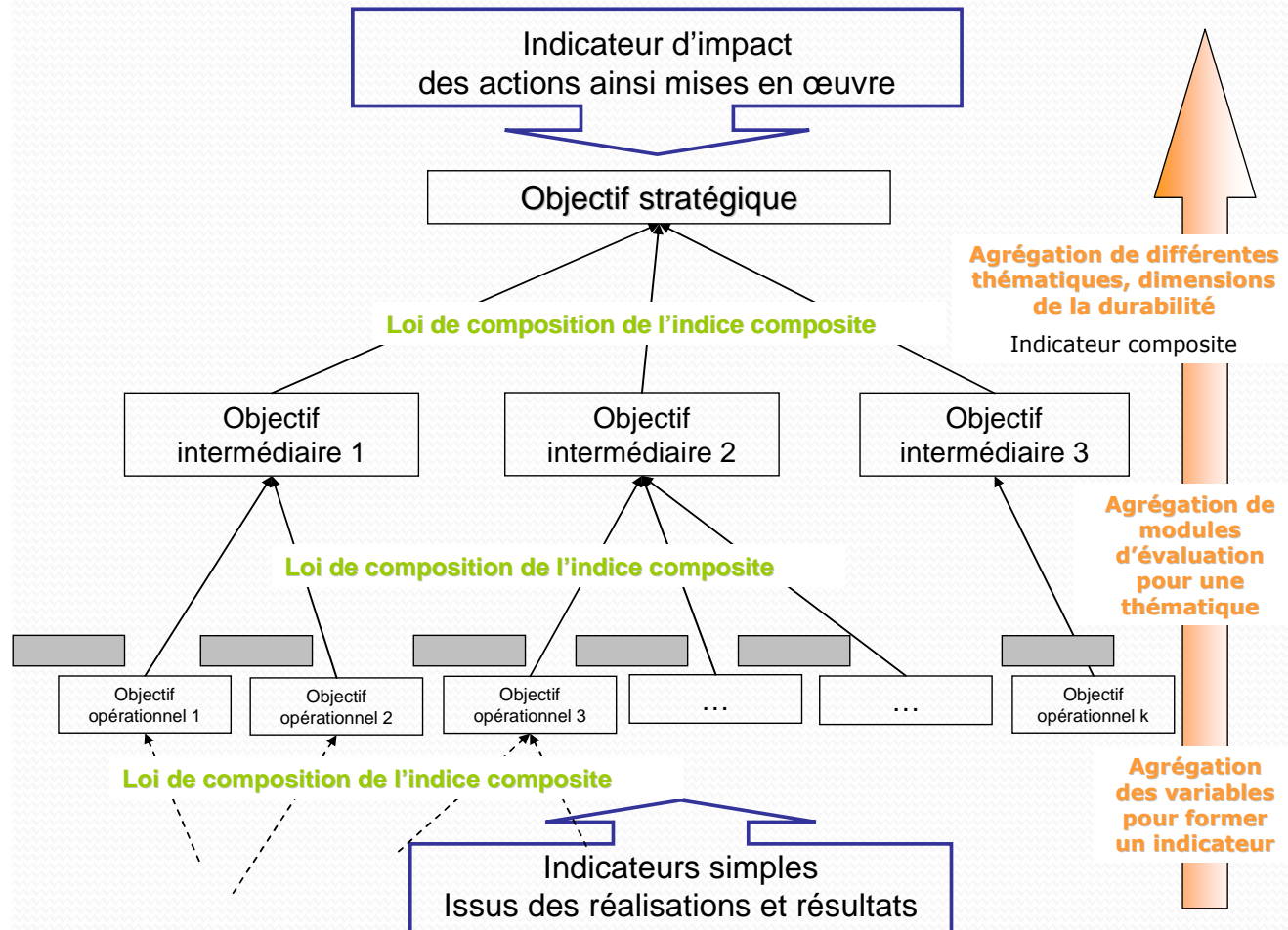
Surtout, ne pas vouloir trop vite la synthèse, ne pas aller vers un classement unicritère.
Comparer des profils de résultats dans le temps avant d'agréger et de suivre un indice synthétique

- -suivre des résultats dans plusieurs dimensions de résultats à l'échelle du territoire : ceci est l'esprit aujourd'hui des PPP en développement durable.
- Ces PPP DD sont expérimentés en situation de risque, d'incertitude et quelquefois par application du simple principe de précaution. Il y a plusieurs objectifs (5 critères de valorisation des résultats): contribution à la modération du changement climatique, à l'économie des ressources et à la protection des milieux, à la réduction des circuits de production et de consommation, assurer une meilleure gouvernance territoriale, développer l'humain.

Savoir optimiser en même temps sur cinq résultats sur le même territoire ?

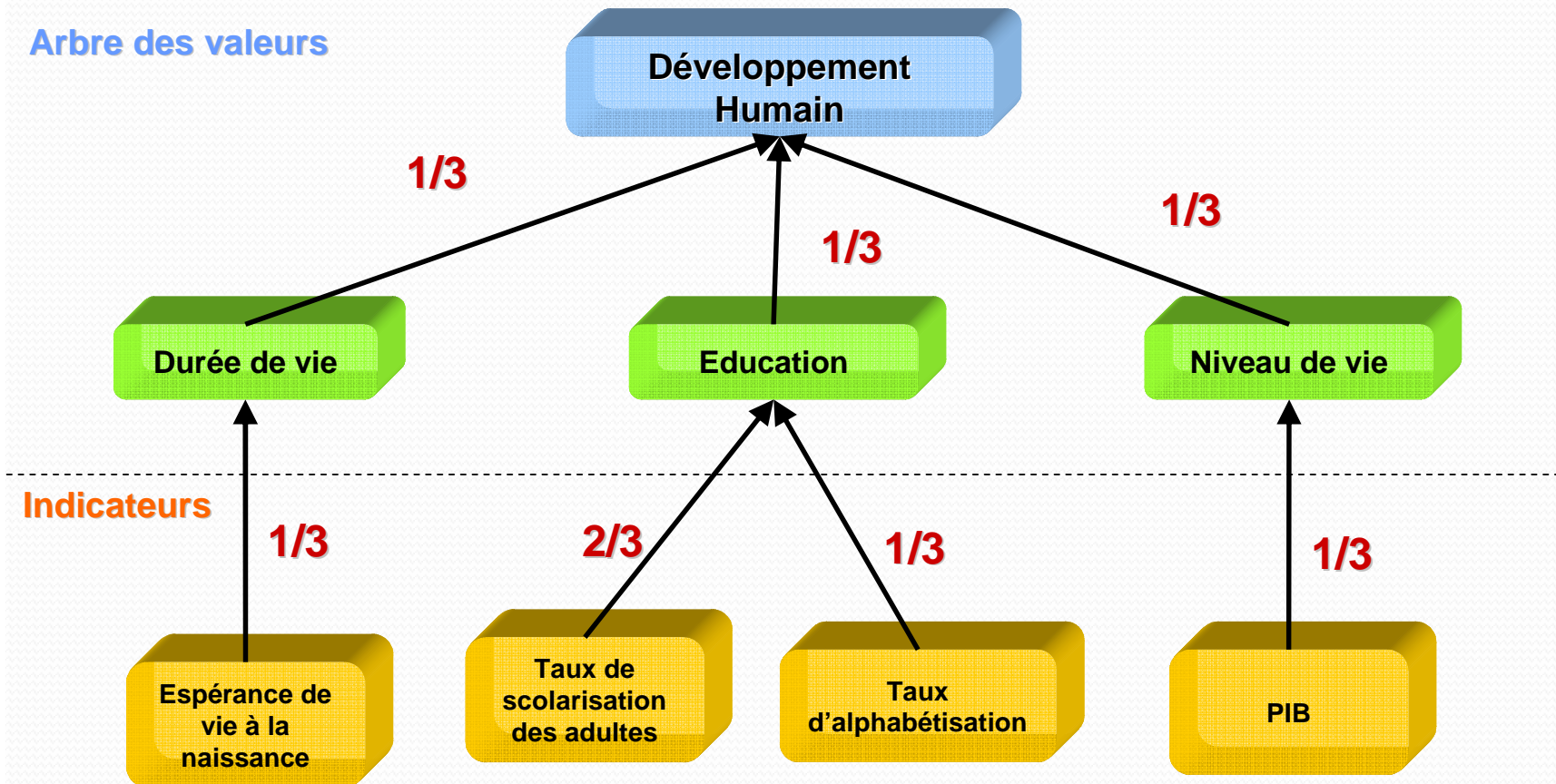


Evitez d'agréger-normaliser les échelles trop rapidement

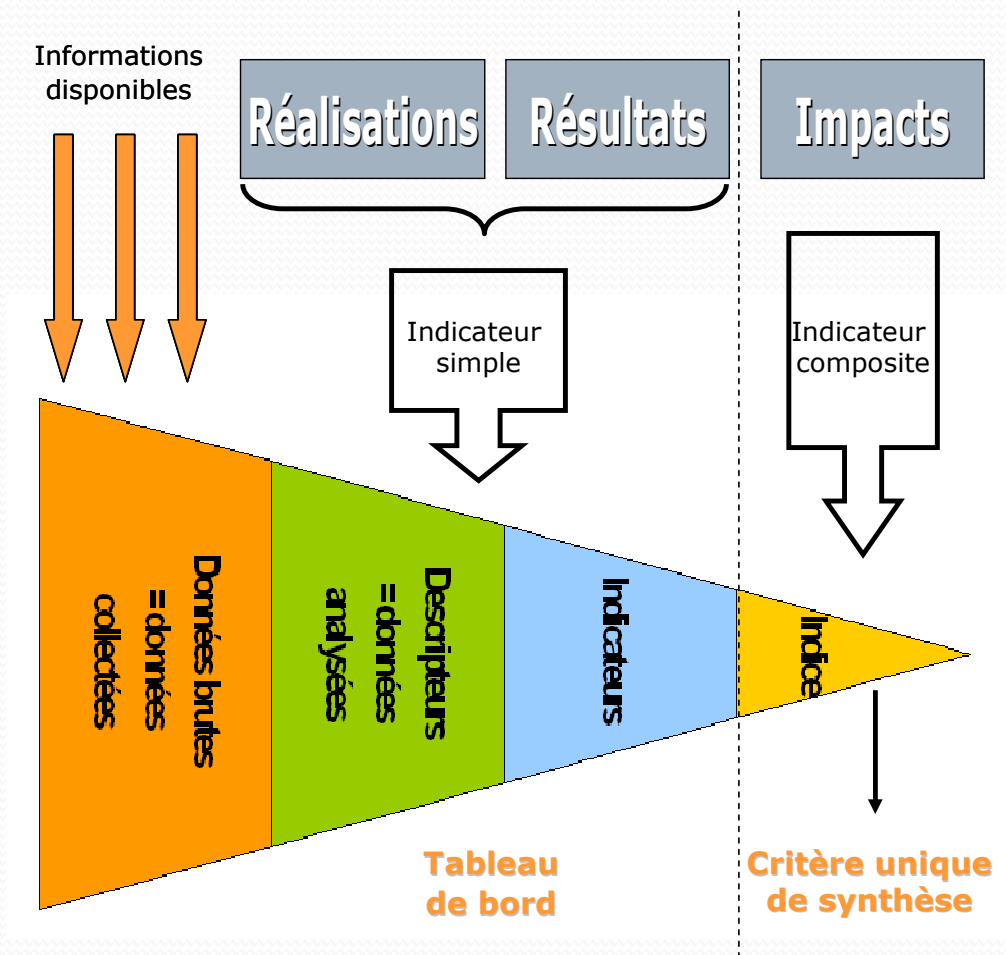


Pourquoi faire une arithmétique linéaire compensatoire des résultats partiels?

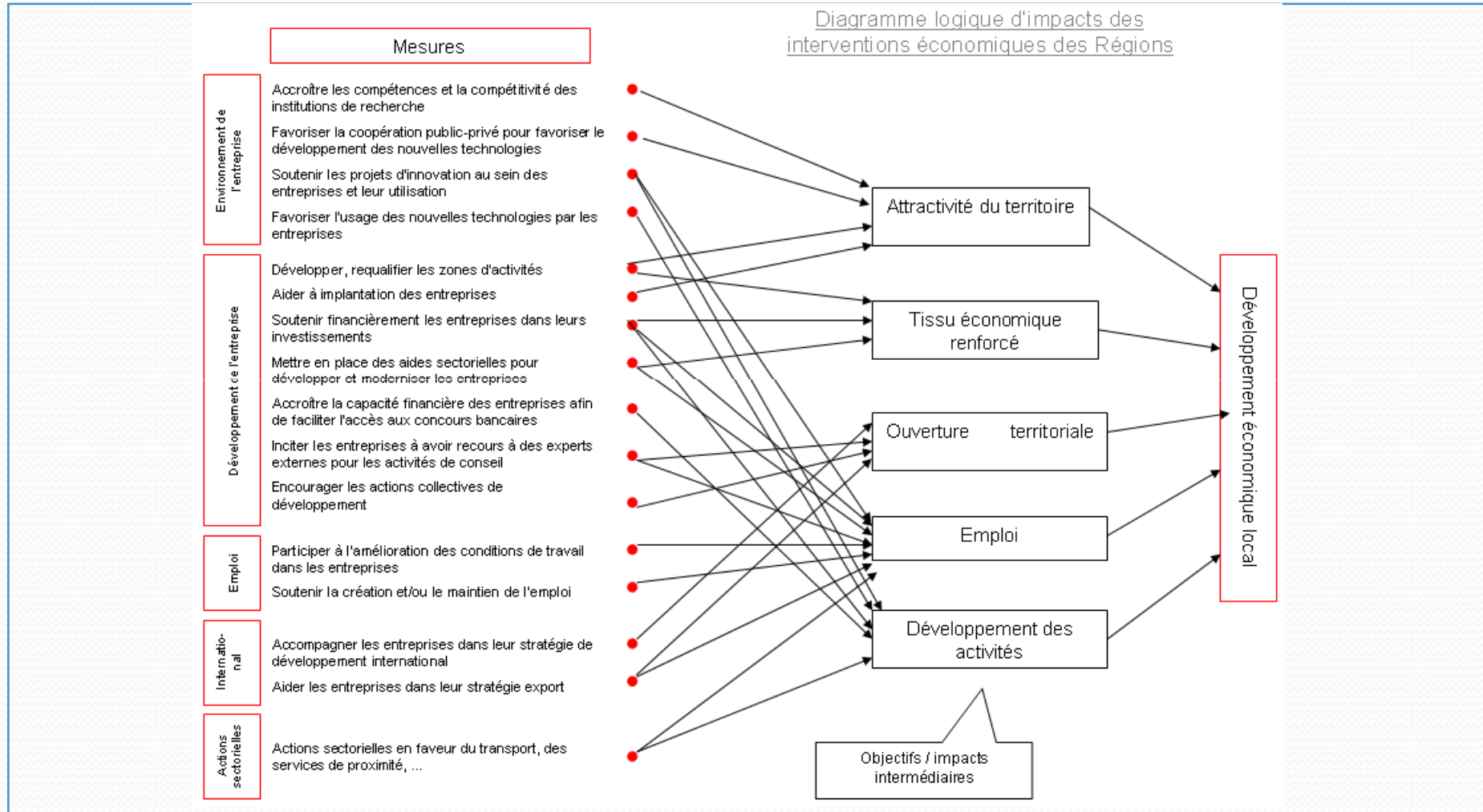
Arbre des valeurs



Il faut identifier les groupes intéressés par chaque type de résultats, ce sont eux qui vont utiliser les résultats des évaluations dans cette dimension de résultat qui leur est propre



Autre exemple : impact sur le développement économique local.



Quitter

Autre exemple : politique universitaire

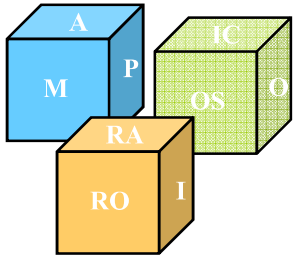


Diagramme logique d'impact



[Nouveau lien](#) [Supprimer lien](#)


Action/Sous-action (A)			Réalizations (RO)			Résultats (RA)			Impacts (I)			
Aides directes	A01	●	●	Organisation de contrôles préventifs dans les écoles et universités	RO3.1	●	●	Amélioration du dépistage préventif de maladie infectieuses	RA3.1	●	Meilleure santé des étudiants et universitaires.	IM01
Aides indirectes	A02	●	●	Campagne d'information	RO3.2	●	●	Amélioration de l'écoute des jeunes	RA3.2	●		
Santé des étudiants et activités associatives, culturelles et sportives	A03	●	●	Actions de sensibilisation dans les soirées étudiantes	RO3.3	●	●	Diminution du ratio des dépenses de santé par étudiant mesuré sur la base des chiffres fournis par les mutuelles étudiantes	RA3.3	●		
Pilotage et animation du programme	A04	●										

3. Il est préférable de recouper les approches et les résultats même avec les évaluations expérimentales!

- Les activités évaluatives doivent faire la « preuve » de l'utilité socio-économique des PPP.
- Il y a différents niveaux de preuve selon les évaluations, selon le cahier des charges, selon la « gouvernance » de l'évaluation, selon la qualité du suivi, selon le caractère plus ou moins approprié de la documentation des réalisations-résultats,
- selon les hypothèses de « force de la preuve » utilisées : hypothèse d'imputabilité directe des résultats aux causes ou selon la simple « probabilité » que l'impact soit le résultat de la combinaison des résultats
- *« Au total, toutes les méthodes d'évaluation ne poursuivent pas les mêmes objets et ne répondent pas aux mêmes interrogations. L'évaluation mobilise un ensemble d'approches à la fois complémentaires et hiérarchisables qui permettent de répondre à des questions différentes »CS HIRSCH.*

Il est préférable de recouper les approches et les résultats même avec les évaluations expérimentales!

- Comme dans tous les raisonnements non purement logiques, une preuve n'est pas une « force probante » définitive.
- Il vaut mieux –modulo le coût des évaluations et le temps qui est un facteur limitant- fabriquer plusieurs preuves : on aura un « faisceau de présomptions »!
- Les preuves doivent être croisées et recoupées « cross-checking ».
- Le rapport final d'évaluation est la démonstration avec l'ensemble des preuves.
- Certains résultats ont une précision faible, d'autres ont une « degré de confiance » faible, ce sont des indices, des présomptions...
-



De toute façon, les résultats ne sont pas repérés sur une échelle unique...etc.
Ceci pose le problème du référentiel des valeurs dans les évaluations.

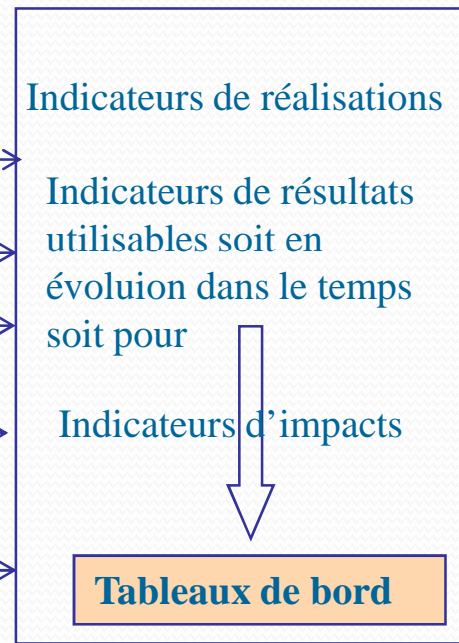
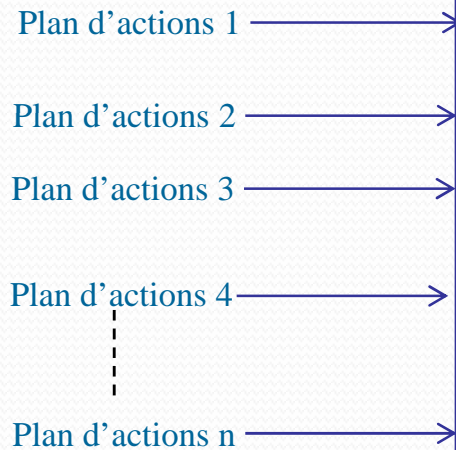
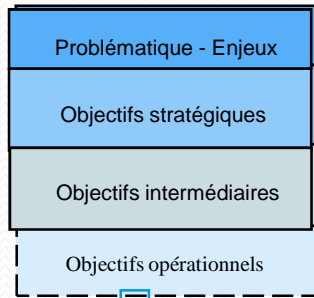
Il peut-être multi-dimensionnel...

Généralement, dans un processus d'aide multicritère à la décision, un décideur désire ordonner les éléments d'un ensemble d'options (résultats par exemple).

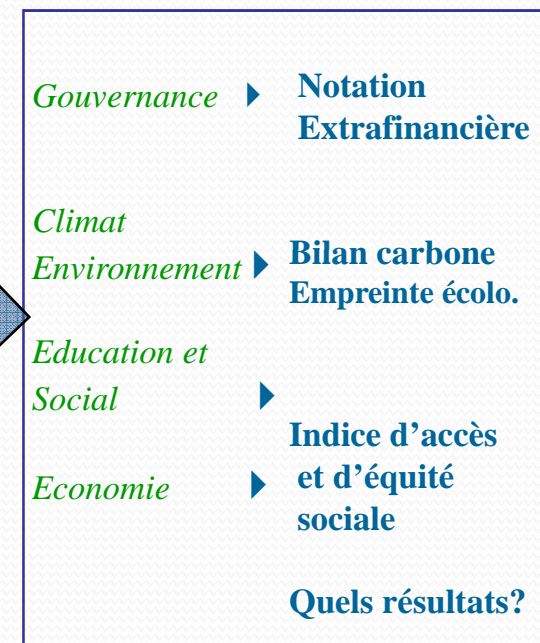
A cette fin, il identifie plusieurs points de vue (juges), dimensions ou critères qui lui paraissent pertinents. Chaque alternative est caractérisée par un profil c'est-à-dire par une performance sur chaque critère. On appelle souvent ceci la matrice d'évaluation ou le tableau des performances.

On va comparer les profils de résultats, les évolutions de profils et essayer de synthétiser.

II - Outils de suivi - évaluation



Mesure des résultats et impacts des objectifs et actions de l'ADP



Indices composites

Baromètre de l'impact de l'action durable du CG

Enjeu des indices composites (ou synthétiques)

Il est préférable de recouper les approches même avec les évaluations expérimentales!

- Les évaluations expérimentales ne peuvent être parfaites.
- Exemple de biais :
- -taux de survie des bénéficiaires, pbs de remplacement des « no-show »
- -pb de « contamination » entre groupe témoin et groupe bénéficiaire (« effet de cafeteria »)

exemple

- Trois méthodes sont préconisées dans le Guide rédigé au nom de la Commission Européenne pour comparer des options de PPP par l'Impact analysis (European Commission, 2008) :
- **L'analyse coûts-bénéfices** qui consiste à identifier et à évaluer les options en fonction de leurs coûts et de leurs bénéfices économiques, sociaux et environnementaux. Une mesure est ainsi considérée justifiée lorsque des bénéfices nets peuvent être retirés de l'intervention publique. Cette méthode ne peut malheureusement pas inclure de données qui ne sont ni quantitatives ni monétaires. Elle est assez longue à mettre en place et coûte relativement cher.
- **L'analyse coûts-effectivité** est une alternative à l'analyse coûts-bénéfices dans le cas où il est difficile d'évaluer les bénéfices en terme monétaire puisqu'elle se concentre sur un seul type de bénéfice. (effets) La contrepartie est qu'elle ne peut pas résoudre le choix d'un niveau optimal de bénéfice et n'apporte aucune aide lorsqu'il s'agit de trouver une option qui conduirait à un gain net pour la société.
- **L'analyse multicritère** enfin regroupe, toujours selon le Guide, un grand nombre de techniques qui partagent le principe de combiner des impacts négatifs et positifs dans un même tableau pour comparer plus facilement des scénarios ou différentes prises de décision. Cette méthode est utile lorsqu'il y a un grand nombre d'informations sur différents impacts ; et que cette information est définie dans différents formats (qualitatif, quantitatif ou monétaire). Cette méthode a pour avantage de reconnaître la multi-dimensionnalité de la soutenabilité et conduit à une présentation relativement claire des résultats.
- Les trois méthodes ont toutes les chances de ne pas ranger les options de la même manière!

Conclusion

- - Il faut avoir une attitude « recherche » dans les activités évaluatives! Il faut « prouver » l'utilité ou la valeur du PPP.
- - Les parties prenantes à la PPP ont intérêt à concevoir ensemble (sociogramme des acteurs ou boule de la gouvernance) et en amont les PPP et leur suivi pour créer la langue du PPP et se mettre en bonne position de suivi (« monitoring ») des réalisations effectives codées par objectif opérationnel (lui-même intégré dans un arbre des objectifs)
- Les cahiers des charges des évaluations devraient conseiller l'utilisation de plusieurs approches, méthodes ou outils..

Conclusion

- Tous les PPP devraient être des expérimentations-évaluations!
- La société fondée sur la connaissance publique des PPP suppose le développement des bases de données de suivi et de bases de connaissances
- Celles-ci devraient être utilisables par les évaluateurs (investissement dans les contenus numériques à l'époque du Web2, cf. Discours PM grand emprunt septembre 2009).

- André S., Roy B. (2006), « Conception et mise en place d'un outil d'évaluation de la performance environnementale », *Journal of Decision Systems*, Volume ?? –
- André S., Roy B. (2006), « Evaluation de la performance environnementale de sites industriels : démarche de concertation pour la mise en place d'un outil d'aide multicritère à la décision », cahier du LAMSADE
- Baslé, Maurice, Huault, Maud (2008). *How to get and analyze Impact composite indexes for European programs and public policy monitoring and evaluations*, in International seminar on impact evaluation of structural funds, Madrid 2008 Session one, New theoretical issues on Impact Evaluation.
- Baslé, Maurice (2008), *Economie, Conseil et Gestion publique. Suivi et évaluation de politiques publiques et de programmes*. Economica, 2008.
- Diact.2009. http://www.territoires.gouv.fr/indicateurs/portail_fr/index_fr.php
- Gadrey J. (2002), « De la croissance au développement : à la recherche d'indicateurs alternatifs »
- Grabisch M. (2006), « L'utilisation de l'intégrale de Choquet en aide multicritère à la décision », *European Working Group Multiple Decision Aiding*, Séries 3, n°14
- Daniel, Kaufmann, Kraay Aart, and Mastruzzi Massimo. "Governance Matters VII: Aggregate and Individual Governance Indicators." *World Bank Policy Research Working Paper*, no. 4654 (2008). <<http://ssrn.com/abstract=1148386>>
- Sandhu-Rojon R. (2003), « Selecting indicators for impact evaluation », PNUD, Partners in Development Evaluation, Paris (http://www.minefi.gouv.fr/fonds_documentaire/TRESOR/cicid/atelier/contrib/29.pdf)
- European Commission (2005), « Impact Assessment Guidelines », SEC (2005) 791. (http://ec.europa.eu/governance/impact/docs/SEC2005_791_IA%20guidelines_annexes.pdf)
- European Commission (2008), « Impact Assessment Guidelines », XXXX 2008, Revised annexes – draft version 27 May 2008, p. 59 (http://ec.europa.eu/governance/impact/consultation/docs/ia_guidelines_draft_annexes_final_en.pdf)
- ONU (Organisation des Nations-Unis), Action 21, Chapitre 28, « Initiatives des collectivités locales a l'appui d'action 21 » (<http://www.un.org/french/ga/special/sids/agenda21/action28.htm>)