



HAL
open science

Schéma directeur pour les données de l'environnement de la Corse

V. Douzal, P. Lemoisson

► **To cite this version:**

V. Douzal, P. Lemoisson. Schéma directeur pour les données de l'environnement de la Corse. [Rapport Technique] irstea. 2009, pp.33. hal-02596489

HAL Id: hal-02596489

<https://hal.inrae.fr/hal-02596489v1>

Submitted on 15 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



territoires, environne-
ment, télédétection
et information spatiale

vincent douzal

Philippe Lemoisson

+33 04 67 54 87 59

+33 04 67 59 37 42

vincent.douzal@teledetection.fr

philippe.lemoisson@cirad.fr

références à rappeler :

vos références :

pièces jointes :

imprimé le :

Montpellier,

027.6 SddeC

schéma directeur pour les données de l'environnement de la Corse

vendredi 19 juin 2009 12:21:00

le vendredi 24 avril 2009

M. Jacques Nicolau
Dreal Corse

copies à

note de synthèse :

schéma directeur pour les données de l'environnement de la Corse

Il nous paraît important de situer le schéma en place en Corse, parmi l'ensemble des dispositifs contemporains.

–1 Une solution très avancée, et opérationnelle dans un paysage complexe, fortement contraint par les standards

Les exigences d'une gestion environnementale durable, et leur traduction en de nouvelles dispositions réglementaires, conduisent les acteurs de la gestion des territoires à organiser un large accès du grand public aux données d'environnement. C'est dans cette perspective qu'une démarche et un dispositif ont été mis en place de façon itérative par la Diren, devenue Dreal de Corse, en élargissant progressivement le cercle d'acteurs, fournisseurs et consommateurs de données en partant de ses partenaires les plus proches. À l'obligation réglementaire de diffuser publiquement l'information, répond le besoin des services d'en disposer opérationnellement : la nécessité de mutualiser les données et réunir les moyens des multiples services en un ensemble fonctionnel ne se discute plus. Les deux peuvent et doivent être satisfaites par un seul et même dispositif.

Une organisation qui ne sera pas un nœud vivant de circulation de données, et notamment de celles qu'elle produit, et qui fera le choix de thésau-

Maison de la télédétection
500 rue Jean-François Breton
F-34093 Montpellier cedex 5

+33 04 67 54 87 54

+33 04 67 54 87 00

www.teledetection.fr

www.cemagref.fr

page 1 sur 8

riser, sera marginalisée. Incidemment, elle sera aussi au ban des exigences réglementaires.

Le dispositif mis en place à Ajaccio est à considérer dans un paysage riche et diversifié de systèmes, portails et logiciels, tous appelés à converger vers des formats et des protocoles d'échange internationalement standardisés. Les processus d'évolution concomitants, lents et délicats, exigent d'importants efforts de concertation, comparaison, confrontation, mise en phase.

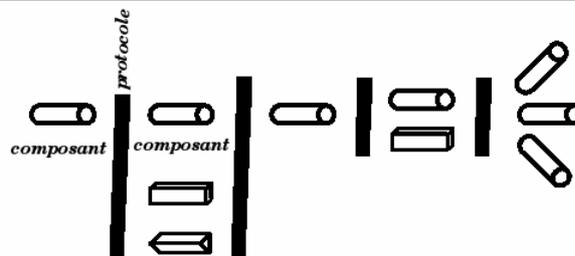


Figure 1 — Tous les systèmes homologues à celui mis en place en Corse visent à se conformer à un même jeu de standards. Ces systèmes sont alors scindés en une série de **composants** fonctionnellement indépendants, qui néanmoins fonctionnent de façon cohérente en se procurant les uns les autres des **services** mutuels, en communiquant selon des **protocoles** strictement, internationalement standardisés. Grâce à l'universalité de l'accord sur ces protocoles, des composants d'origines diverses communiquent entre eux et collaborent fluidement, ils ignorent même l'identité des composants avec lesquels ils échangent des services, ils savent seulement qu'il se conforment au même langage d'échange. Les composants qui partagent un même protocole sont dits **interopérables**. Le composant demandeur est classiquement baptisé **client**, celui qui lui répond est un **serveur**. L'architecture générale résultante est dite en **couches** (tiers en anglais, c'est-à-dire étages), bien qu'elle puisse se ramifier, et plus encore que ne le suggère la figure

Cette architecture plébiscite une morale établie par une cinquantaine d'années de travaux et d'essais en informatique, à savoir que dans un système, « **les données dominent** » : leur organisation, leur structuration détermine ce qu'il est possible d'obtenir du système. Toute cette configuration en couches est effectivement commandée par les structures de données échangées, **présentées** dans des langages d'échange, les protocoles.

Le mouvement de standardisation est parti des protocoles les plus fondamentaux d'échanges et de routage sur les réseaux, et en vient maintenant à s'intéresser au détail **sémantique** des données que s'échangent des services hautement spécialisés dans un **métier** donné. On note par exemple les efforts de normalisation du vocabulaire sur l'eau conduits en France par l'Onema (Office national de l'eau et des milieux aquatiques) et mis en œuvre dans le Sandre (service d'administration nationale des données et référentiels sur l'eau). Ce processus, lent et délicat, puisqu'il peut imposer le démontage d'applications à architecture monolithique, exige d'importants efforts de concertation, comparaison, confrontation, mise en phase.

Certains systèmes sont relativement peu avancés dans ce mouvement migratoire. Par exemple, Cartorisque présente au public un panorama général des données et cartes sur les risques en France ; cependant à ce jour, l'alimentation des données est asynchrone, les producteurs de données transmettent épisodiquement les lots et la consultation passe souvent par un téléchargement préalable qui rompt la **chaîne des services**. Cette situation en fait essentiellement un stock de données mortes, une sorte de cul-de-sac : superficiellement, on obtient des indications, mais dès qu'on a un besoin précis, il excède ce qu'on peut obtenir des données disponibles.

C'est parce qu'un ensemble de protocoles instaurant un schéma de découpe en composants qui assurent des services déterminés est en place aujourd'hui et continue à se développer, que toute tentative de solution qui ne respecterait pas ces normes d'interopérabilité serait aujourd'hui vouée à l'échec. Grâce à cette standardisation, un ministère peut investir dans le développement de logiciels qui réalisent un tel ensemble de composants sans être à la merci d'une offre commerciale exclusive — c'est ce qu'a fait

celui en charge de l'écologie. Il a la certitude de la pérennité de son investissement, de ne jamais voir ses solutions marginalisées, ni son système totalement remis en question. Qui plus est, le fait qu'il rende le code de ses logiciels libre contribue à asseoir la vigueur des standards adoptés et garantit aussi la durabilité de l'investissement.

Cependant, à ce jour, on est encore dans **un vaste mouvement de migration** simultanée des systèmes d'information et des offres logicielles. D'une part la multitude de systèmes d'information, que tous les types d'organisations publiques ou privées entretiennent, évolue vers ces architectures de services. D'autre part les offres logicielles elles-mêmes migrent vers une conformation complète à l'immense palette de protocoles qui sont établis ou en cours de définition. Ces mutations simultanées rendent certaines analyses délicates.

Conformément aux considérations de la figure 1, le dispositif Corse doit s'analyser en **deux composantes** : d'une part, **des données**, présentées sous forme de services standard (WMS, WFS, bases SQL, etc.), ou, au minimum, de formats standards, dont la description est publique. D'autre part, **une solution logicielle**, comprenant un outil de cartographie dynamique et un catalogue, qui permet de suppléer au manque d'équipement d'un utilisateur (un « client » des services) en logiciel de système d'information géographique (SIG) par exemple. Il lui suffit alors d'un ordinateur muni d'un navigateur web pour avoir un plein accès aux données, y compris cartographiques, ce qui, si on lui en confère le droit, inclut la capacité à les modifier.

La solution logicielle en place en Corse est dans la conformité la plus complète qui soit à ce jour avec les standards concernant les échanges de données, en particulier géographiques. Relativement à d'autres offres¹, elle possède quelques avancées techniques, par exemple la capacité à procurer un véritable SIG en ligne, avec l'accès aux bases de données sous-jacentes, aux entités géographiques, mais avec aussi la possibilité de les mettre à jour, ce qui permet de répartir les personnes qui font ce travail souplement sur un territoire quelconque, pourvu que des moyens de communication existent.

La solution logicielle qui sous-tend le dispositif mis en place à Ajaccio est un produit du commerce, ARCGIS d'Esri, en appui sur une base SQL server. Cette solution étant conforme aux standards établis, elle permettrait d'envisager une mutation vers des solutions libres, ou autres². Cependant, une telle mutation semble ne présenter aucun intérêt à un horizon visible à ce jour. D'une part, l'investissement en licences est fait, et les charges de maintenance locale ne seraient pas supprimées pour autant. D'autre part, la solution est aujourd'hui pleinement opérationnelle, et les services attendus par les utilisateurs finals sont rendus, ce qui est la raison d'être du dispositif³. Migrer vers un autre dispositif logiciel serait un travail important, qui immobiliserait l'équipe et tous les partenaires pendant une période sensible, pour aboutir au mieux au même niveau de fonctionnalité final (et, comme on l'a noté plus haut, probablement à un niveau de fonctionnalité inférieur).

¹ Par exemple l'offre Carmen, supportée par le ministère en charge de l'écologie, comparaison faite à la date de la rédaction.

² Une telle migration serait d'autant plus facilement réalisable que les idiosyncrasies qu'offrent ces logiciels, qui sont autant de pièges non-standards, ont été attentivement et soigneusement évitées par le concepteur lors de l'implémentation. Ainsi, le dictionnaire des données de l'infrastructure Corse est-il enregistré dans une table standard. Les créateurs de logiciels commerciaux ne manquent jamais, tout en se tenant aux standards établis par le marché, d'offrir des « améliorations » par lesquelles ils se donnent des chances de garder leurs clients captifs. De même les « géodatabases », de format privé ESRI, ne sont utilisées que comme des contenants pratiques pour des lots de données de référence.

³ Ils sont même rendus, ou tout au moins offerts, bien au-delà de ce qui est effectivement entré en usage au près des utilisateurs, qui pour y venir pleinement doivent ajuster leur propre système pour qu'il se prête mieux à une architecture ouverte, en services. Ces ajustements, même mineurs, interagissent avec les politiques propres de chaque service qui ne sont pas encore conduites en harmonie, spécialement s'agissant des services déconcentrés de ministères divers, d'où un gros travail de mise en phase et discussion.

Or, comme on y reviendra au§-4, le cœur du travail se situe ailleurs, dans les données elles-mêmes.

–2 Une grille de cinq critères pour évaluer des systèmes apparentés au schéma Corse

Nous proposons de dégager cinq critères permettant de fonder une éventuelle comparaison entre des « dispositifs » dédiés au partage d'informations dans un contexte où les sources comme les usages sont multiples. Ces critères se déploient dans des dimensions largement indépendantes, qui néanmoins se recoupent et se connectent ; l'ordre dans lequel ils sont présentés ci-dessous n'est pas significatif :

1. **consensus autour de la signification des données** : les données ne deviennent des informations que dans la mesure où elles font sens pour des acteurs. La question délicate du partage du sens se met en place au cours d'un processus qui fait partie de l'activité d'**administration des données**. Évaluer un dispositif sous cet angle passe par la réponse aux questions suivantes :

- y a-t-il un catalogue de métadonnées ?
- y a-t-il un dictionnaire des données stockées ?
- y a-t-il un vrai index ?
- leur processus de construction a-t-il bien impliqué tous les acteurs, et
- y a-t-on bien intégré et décrit leurs **schémas d'usage** des données, qui ultimement seuls en certifient la signification (sinon, le sens des données est enfoui dans le formalisme des descriptions de structures de données, qui le rendent hermétique.) ?

2. **propriété, contrôle, responsabilité vis-à-vis des données** : il est communément admis aujourd'hui que la donnée doit rester « au plus près de son producteur » ; ce qui signifie qu'il a tout contrôle sur leur publication et leur diffusion. C'est une traduction du principe de **subsidiarité**. L'idée de proximité doit être traduite dans un sens fonctionnel : le disque où sont enregistrées les données n'a pas à être à proximité de l'opérateur, mais celui-ci doit avoir toute latitude pour les manipuler, et la maîtrise complète de leur service de mise à disposition.

Le **droit** qui rend désormais obligatoire l'instauration de ces dispositifs pour rendre les données publiques vient se superposer à **d'autres régimes juridiques** qui lui placent des limites, par exemple, le code civil qui pose le respect de la vie privée. On peut s'interroger sur les responsabilités qui peuvent découler de la publication de données. Mais la diversité des acteurs et des situations dans lesquelles peut survenir un dommage qui se raccorde au droit de l'environnement, et peut donc par exemple mettre en cause des données de l'environnement est telle, avec une telle imbrication potentielle de régimes juridiques, qu'il est douteux de présumer en prévoir l'étendue exacte.

Par ailleurs, on se trouve dans une situation de **régime juridique non établi**, une situation inédite, dont le droit propre n'est pas constitué, puisqu'il croise deux domaines eux-mêmes nouveaux, à savoir la transparence des données de l'environnement et la disponibilité de vastes jeux de données, avec des facilités de manipulation inédites, grâce à leur forme électronique. La possibilité d'intégrer, de combiner des données (la vitesse, la facilité avec laquelle on le fait) fait exploser la combinatoire des assemblages, des recoupements qui peuvent être produits chaque jour, combinatoire elle-même amplifiée par de nouvelles modalités de croisements. L'utilisateur naïf peut se croire fondé à opérer tous les assemblages techniquement possibles. Mais de même que toutes les phrases grammaticalement correctes de la langue française ne sont pas sensées, toutes ces combinaisons n'ont pas nécessairement légitimes. On ne peut ni définir par avance ce qu'il est valide de faire, ni énumérer tout ce qui est exclu. Comme dans toute situation similaire, face aux inquiétudes que peut susciter ce vide, il n'existe pas de meilleure protection que de faire platement état de la façon dont sont obtenues les données, et l'**usage** auquel elles sont desti-

nées, qui est celui pour lequel elles ont été testées, donc validées. C'est d'ailleurs ce que demanderait un juge qui aurait à régler un litige.

Des questions essentielles sont donc :

- les questions de droit sont-elles abordées dans le processus de mutualisation ?
- Est-ce que ces questions sont proprement arrimées aux processus d'acquisition des données, à l'**usage** qui en est fait, en documentant soigneusement ces deux facettes ?
- Est-ce qu'un avertissement liminaire, et un texte adéquat plus développé permet aux utilisateurs de prendre la mesure des questions soulevées par la manipulation de données, en l'incitant à se plonger dans la documentation des données ?

3. **qualité des données** : cette notion n'est pas toujours facile à appréhender, la non-qualité pouvant provenir non seulement d'erreurs de mesure ou de saisie mais également (et peut-être surtout) d'une ambiguïté relative à l'**usage** qui peut être fait de l'information (point 1.). La meilleure méthode pour se prémunir de l'erreur est la « boucle de rétroaction » selon laquelle le producteur de la donnée est soit son premier utilisateur, soit en relation directe avec ses utilisateurs. La meilleure méthode pour se prémunir de l'ambiguïté est la description précise de la façon dont la donnée a été mesurée (la chaîne d'acquisition des données). C'est la meilleure qualification, la meilleure documentation que l'on puisse apporter en vue de sceller le point 1. Ainsi à la qualité des données, substitue-t-on une **qualification des données**. Les données, donc, sont-elles pleinement qualifiées ?

4. **accessibilité des données** : il ne suffit pas que des données de qualité, dont la signification fait l'objet d'un consensus, existent dans un dispositif ; encore faut-il que tous ceux qui en sont en droit d'en avoir l'usage puissent y accéder et les manipuler sans devoir souscrire à une configuration particulière (autre que de se conformer aux standards en vigueur). Un dispositif fournit des données techniquement accessibles quand il obéit aux standards, normes techniques d'interopérabilité.

5. **pérennité des données** : cette dimension est trop souvent laissée pour compte et pourtant son importance est fondamentale, particulièrement lorsque les données gérées doivent permettre d'observer et décrire des évolutions, de mesurer des tendances à long terme, de servir de référence en cas de litige ultérieur. Il faut assurer la pérennité de l'ensemble de la chaîne technico-logicielle qui permet de lire les données, de façon à respecter, maintenant que les données sont exclusivement sous forme numérique, les principes présidant à la loi sur les archives classiques. Évaluer un dispositif sous cet angle passe par la réponse aux questions suivantes :

- y a-t-il uniquement des formats standards de stockage basé sur des normes ouvertes ?
- y a-t-il sécurisation des serveurs, réplication des données, sur des sites distincts et distants, sauvegardes régulières ?
- l'organisation des bases, la structure du système permet-elle de conserver et d'étudier des chroniques, y compris à long terme (ce qui protège aussi de pertes majeures) ?

Cette grille est un support de lecture, pour quiconque déchiffre un système d'information. Elle est présente en permanence à l'esprit de quelqu'un qui s'implique dans la conception, l'animation, la maintenance d'un système, spécialement un système mutualisé.

— **Il apparaît que la question de l'usage des données est le fait organisateur du système.** Largement développée dans le rapport 025. SddeC, elle est soulevée, sous une forme ou une autre, sur les cinq points d'analyse, et, quand on observe un système, sur ses aspects les plus abstraits comme ceux les plus pratiques. Par exemple, la prise en compte des conditions d'usage effectif, du travail concret effectué par les opérateurs qui nourrissent le jeu de données est essentielle pour assurer un système efficace, qui incite à maintenir des données de qualité, à jour. L'absence d'usage, ou la déconnexion entre l'acquisition et l'usage permet

d'identifier des pans de données sans qualité, sans qualification, on parle couramment de cimetières de données.

— Concrètement, **l'avertissement sur l'usage des données** figurant sur l'atlas des données de l'environnement de Corse est, à ce stade, une information adéquate ; seule la mention stipulant que les données sont fournies à titre informatif, sans valeur réglementaire doit être amendée, puisqu'elles ont une certaine valeur. Cependant, il faut préciser qu'elles ne sauraient se substituer aux procédures d'information ou d'autorisation, comme celles du fait de la loi sur l'eau, ou les permis de construire. **La nouveauté de cette situation de droit incite à en faire l'objet d'une étude spécifique.**

–3 De cette grille à une comparaison approfondie des systèmes

On pourrait être tenté, sur la base de cette grille d'analyse, de comparer entre eux différents dispositifs, et notamment un certain nombre de ceux qui existent, survivent ou se développent en France (Carmen, Prodige, le Géoweb et le Géorépertoire du ministère de l'agriculture, Cartorisque, Cartélie, etc.). Mais conduire une telle analyse supposerait en premier lieu un accès à ces dispositifs pour pouvoir les appréhender de l'intérieur, et en second lieu de pratiquer de façon approfondie les manipulations courantes d'alimentation en données, d'administration, d'usage. Il s'agit évidemment d'un travail conséquent, qui serait par ailleurs extrêmement péremptible car tous ces systèmes sans exception, s'ils ne sont pas en cours d'abandon, sont en pleine migration vers une structure plus achevée en services telle qu'évoquée plus haut.

Accéder à l'intérieur de ces systèmes semble aller de soi, s'ils répondent aux principes généraux qui sont réalisés en Corse. Mais les niveaux de migration sont très différents. Prodige, par exemple, n'est absolument pas évaluable du point de vue d'un citoyen utilisateur, car il n'est accessible, à ce jour, même en simple consultation, qu'aux services de l'État, et parmi eux, à un sous-ensemble restreint de personnes accréditées. C'est là que la bât blesse : clos, il ne répond pas aux exigences des textes réglementaires sur l'accès à l'information cités dans le rapport 025.SddeC, il est du même coup hermétique à une exploration et a fortiori une évaluation publique.

L'impossibilité de cette comparaison détaillée n'empêche pas de se former une idée opérante sur un système. Il faut adopter un mode de raisonnement où l'on admet que l'on ne peut contrôler complètement un processus, ni la direction qui sera retenue à l'avenir, et où l'on contrôle que l'on se trouve dans le bon mouvement. Par ailleurs, il est impossible d'avancer seul, tout comme de laisser un partenaire sur le bord du chemin.

Enfin, il faut se convaincre que quelle que soit la sophistication des technologies et des méthodes employées, rien du système d'information n'existe de façon tangible, en l'absence de données. Cette domination des données achève de dissuader de tenter une comparaison entièrement développée. La comparaison ne pourrait s'arrêter à la surface des principes énoncés. Il y a obligation de plonger dans les modèles conceptuels de données, puis dans le contenu même de ces données, d'examiner à l'aune de nos critères chacune de leurs sources, chacun de leurs usages effectifs qui conditionnent toute la signification qu'elles revêtent. Au final, il ne s'agit pas d'un point de passage, d'une étape à franchir ; c'est **la matière même du schéma**, l'activité quotidienne de l'administration de ses données :

–4 Les données, leur inventaire, leur documentation, concentrent l'essentiel du travail

C'est d'abord une conséquence de ce que les **données dominent**. C'est aussi parce que les données touchent à la matière même des métiers (point 1) et à la façon dont ils s'exercent. Il y a un vaste travail de refonte, de reconception qui accompagne le crible des données. Il faut aller au cœur des données, des usages, des besoins. Et comme les questions que l'on se

pose sur l'environnement sont en renouvellement perpétuel, ce travail sur les données se mutera en une activité entretenue, permanente, de même qu'aucun logiciel ne peut perdurer sans maintenance. Le rapport 025. SddeC parle d'inventaire permanent, à outiller et animer.

Le travail sur les données, qui consiste à s'entendre sur leur sémantique, c'est-à-dire établir des protocoles sur leur sémantique, et à documenter ce processus, est donc à la fois l'urgence, la priorité, et le gros du travail. C'est aussi l'activité essentielle du schéma.

Un tel travail d'inventaire et d'analyse des lots de données est au cœur du processus en cours dans le schéma actuellement en vigueur à Ajaccio. L'évolution continue des besoins, des usages sur les données implique par ailleurs que ce travail devienne un inventaire permanent.

Les partenaires qui ont un besoin mutuel doivent instituer un organe qui concentre et confronte les besoins et les avis de tous, et établit en commun un protocole (c'est-à-dire, quel que soit le sujet, une entente sur un vocabulaire, et la façon de former des phrases valides). Cette nécessité d'unification relève du **principe de suppléance**, l'exact miroir du principe de subsidiarité qui permet, l'entente acquise, à chacun de s'appuyer dessus pour assurer sa part de fonctionnement. C'est un processus spontané, et non pas une autorité classique, qui peut élire un partenaire comme moteur de cette suppléance, selon un processus. tout à fait similaire au fonctionnement du monde du logiciel libre.

Dans le périmètre français, et sur l'eau, on a vu que l'Onema a été instauré, à l'appel d'un ensemble d'acteurs, à partir d'un certain moment, pour assurer ce rôle de suppléance. La suppléance n'épargne à aucun des acteurs un **travail d'implication**, au minimum participer aux travaux communs. Elle ne peut non plus leur économiser l'**effort de spécialisation** ; qui consiste à produire, à proposer les éléments qui correspondent à une compétence qu'il est le seul à détenir, des missions qu'il est le seul à exercer, un usage qui lui est propre, une fois que cette spécificité a été bien délimitée collectivement. Cette délimitation consiste à scinder le domaine de travail en protocoles, articulant des composants (figure 1).

Pour ce qui est des données de l'environnement en Corse, la Dreal joue naturellement ce rôle de moteur dans l'animation de la suppléance, à la suite de l'action de la Diren.

Le rapport 025. SddeC a été écrit pour permettre de prendre plus substantiellement la mesure des enjeux et du travail que représente l'établissement d'un dispositif comme le schéma directeur des données de l'environnement de Corse. Il développe et articule entre eux les principes auxquels on ne peut déroger pour publier des données et propose des éléments pour accompagner l'évolution de l'administration des données vers une configuration qui réponde à ces principes. Enfin, dans son contenu même, il propose un point d'entrée pour permettre à un honnête homme utilisateur, intéressé par les questions d'environnement mais pas au fait des arcanes des technologies modernes, de se former une vision, un cadre mental qui lui permette de naviguer à la recherche d'informations. Des comptes rendus circonstanciés des discussions qui président à l'établissement de consensus sur la signification des données, sont aussi une irremplaçable information pour permettre à un nouvel arrivant, par exemple un utilisateur qui découvre le système, de comprendre sans ambiguïté le vocabulaire qui est employé, les données qu'il va manipuler.

-5 Conclusion

Le schéma a adopté les protocoles d'échange et les principes d'organisation incontournables, et les a effectivement mis en œuvre et amenés à un point de maturité avancée. Techniquement, les éléments fondamentaux sont en place, et c'est déjà un accomplissement dans le paysage contemporain. La Diren a été motrice des processus.

Il incombe à la Dreal de poursuivre et d'entretenir le travail de fond du §-4, en continuant à entraîner ses partenaires vers un ensemble opérationnel de « services de données » répondant aux critères énoncés plus haut.

La clef du succès de l'entreprise réside dans le travail intime sur les données. Le caractère pionnier de l'entreprise implique un investissement important et donne une responsabilité toute particulière à celui qui aura eu le mérite de tracer le premier sillon. De plus, la multiplicité des thèmes, la spécificité du territoire où cet assemblage a lieu impliquent un travail original de spécialisation, toujours articulé avec les autres forces normalisatrices emboîtées, nationales, européennes, internationales. Le travail déjà effectué et l'effet d'entraînement qu'il produit le placent sous les meilleurs auspices.

vincent douzal Philippe Lemoisson



territoires,
environnement,
télédétection
et information spatiale

vincent douzal
Philippe Lemoisson

références à rappeler :
vos références :
pièces jointes :
imprimé le :

Montpellier,

+33 04 67 54 87 59

vincent.douzal@teledetection.fr
philippe.lemoisson@cirad.fr

025.3 SddeC

vendredi 19 juin 2009 à 12:12:00

le lundi 2 février 2009

M. Jacques Nicolau
Diren Corse

copies à

rapport d'étape, pour le

schéma directeur des données de l'environnement de Corse

-1 Introduction

Le but de ce texte est d'introduire l'ensemble des problèmes du « **schéma** », comme on l'appellera simplement dans la suite, dans un cadre de lecture global, intelligible par un très large public. Cette présentation permet de dégager un **ensemble de principes** qui doivent présider à sa mise en œuvre.

Le schéma soulève un grand nombre de questions qui ne peuvent être abordées qu'en ayant intégré un bagage technique très important, qui pousse des tentacules dans des nombreux domaines habituellement déconnectés, et qui n'existe nulle part regroupé en un ouvrage unique auquel on pourrait se reporter. Ces questions doivent devenir accessibles, palpables, pour pouvoir devenir des objets de débat, pour la très large palette de personnes appelées à intervenir dans le cadre du schéma — et au moins pour commencer, les acteurs primaires, membres de toutes les organisations compétentes à un titre ou un autre sur les questions d'environnement en Corse.

Nombre de personnes cultivent un complexe par ignorance devant la technique et le vocabulaire informatique qui déferlent désormais sans prévenir sur à peu près tous les métiers. Pourtant, un système d'information peut être parfaitement réglé avec du papier et des crayons, il

Maison de la télédétection
500 rue Jean-François Breton
F-34093 Montpellier cedex 5

+33 04 67 54 87 54

+33 04 67 54 87 00

www.teledetection.fr

www.cemagref.fr

page 1 sur 25

peut aussi mobiliser la multitude de sophistications qu'offre la technologie informatique et être en discordance totale avec la réalité. Au milieu de la tempête de concepts informatiques, on perd les principes de jugement, les concepts de bon sens que chacun possède bien à sa façon, et qui valent quel que soit le mode technologique du système d'information. Ce document sera peut-être une aide à la navigation dans ces mers agitées.

Le schéma vise premièrement à permettre à tous les services des acteurs primaires de **travailler conjointement** sur leurs données.

C'est un **processus** en cours, établi sur des principes bien posés et un processus d'agrégation des partenaires qui fonctionne.

Deuxièmement, le schéma doit permettre au public d'**accéder** aussi pleinement que possible aux données sur l'environnement.

–2 Le sens de l'histoire et le développement durable : publier sur l'environnement, sur le fonctionnement des milieux

–2.1 . L'invention des données est synchrone de l'invention de l'écriture

Depuis l'invention de l'écriture sur des tablettes d'argile à Sumer, on a procédé à des inscriptions sur des **mémoires externes** dans le but de fixer la teneur d'un **contrat**, d'assurer la gestion d'une transaction, une convention impliquant deux **protagonistes** dans le cas le plus simple, ou toute une collectivité autour d'un **bien essentiel** dont il fallait régler l'usage commun : tenir des comptes, assurer la perception de taxes, etc.

Les **données** fixent la teneur d'un contrat, il est inscrit en une **représentation symbolique** sur une mémoire externe qui rend **public** le contenu du contrat à tous ceux qui sont amenés à le consulter, à lire la mémoire. L'écriture, le contrat fonctionne selon un **schéma de communication**, c'est-à-dire un **jeu de conventions** partagées entre les protagonistes pour assigner une relation entre la connaissance d'un état du monde et son enregistrement sur une mémoire externe. On remarque aussi, au passage, que ce jeu de conventions qui permet l'écriture est déjà une forme de contrat, sous-jacent au premier (contrats et conventions, dans l'usage juridique, sont de quasi-synonymes). On voit donc esquissé un fonctionnement en **couches successives** empilées, jouant chacune pour celle qui la surmonte le rôle que lui assure la couche sous-jacente. Deux parties conversent selon les termes « de haut niveau » d'un contrat, parce qu'au niveau inférieur, les écritures employées de chaque côté sont concordantes, en quelque sorte elles aussi conversent entre elles. On retrouve cette structure dans les couches Osi pour la transmission d'information sur les réseaux informatiques, qui posent les principes à la base d'Internet et du web.

Deux couches homologues, quel que soit leur niveau, situées aux deux extrémités d'un canal de communication y conversent selon un protocole déterminé. Elles acheminent chacune leurs message en le transmettant simplement au travers de l'interface avec la couche sur laquelle elles reposent. Ainsi un serveur web et un navigateur échangent des messages selon le protocole http, comme s'ils se parlaient directement, ces messages sont transférés notamment aux couches de transport, puis aux couches physiques qui assurent la transformation du codage qu'on leur transmet en impulsions électriques, lumineuses ou en ondes radio, et la restitution du code à l'autre extrémité.

L'assignation d'un état du monde à un état de la mémoire externe, l'établissement d'un rapport entre le monde et une formalisation qu'on en fait est ce qu'on appelle une opération de **mesurage**, dans un sens très général. C'est par elle que l'on obtient une mesure, c'est-à-dire un **symbole** d'un **alphabet** ou **code**, à partir d'une configuration donnée du monde. L'opération d'assignation demande, de la part de l'individu qui fait la mesure, l'intervention d'une **symbolisation**, une catégorisation, par

laquelle il **juge** que ce qu'il perçoit est représenté adéquatement par un symbole donné dans le modèle qu'il se fait du monde.

Le **code de symboles** est un support objectif pour tous ceux qui savent le lire, il forme un **langage d'échange**. On peut à tout moment se rapporter au support de mémoire externe pour vérifier ce qui y a été inscrit, autant qu'il est inviolable, inaltérable. Le code est parfaitement déterminant, sous couvert des **interprétations** aux deux extrémités de la chaîne de communication, par lesquelles on transforme la perception d'un état du monde en un code (**mesure**) ou un code en une **action** dans le monde. Ce sont les symbolisations, que seuls des hommes peuvent effectuer.

Transportable, la mémoire externe qui porte ce résultat de mesure sert de média, elle permet de dépasser la portée d'un contrat oral, qui suppose un contact direct, face-à-face réitéré des parties. On peut ainsi déporter la satisfaction du contrat auprès d'un tiers, déplacer le contrat : c'est une donnée.

Des données, quelles qu'elles soient, entrent toujours dans la structure d'un schéma de communication, l'encodage est un langage, c'est un moyen de **publier un relevé convenu** d'un état de la réalité. Dans le cas le plus simple, deux personnes seulement sont concernées, mais le schéma peut impliquer de diverses façons de multiples intervenants qui vont donc communiquer de diverses façons.

Quand on quitte le registre du contrat oral pour celui d'un contrat écrit, c'est-à-dire d'une **base de données**, survient un écueil, un changement immédiat : comme il n'y a plus les liens entretenus de compréhension, d'attachement, de partage, de mise en phase culturelle grâce à la proximité, au contact direct fréquent entre les **parties prenantes**, il y a possibilité qu'un fossé d'incompréhension se creuse, un fossé bien plus vaste que ce que la communication courante directe offre déjà d'occasions de quiproquos. Il y a une décohésion, une déconnexion **sémantique**, c'est-à-dire sur la valeur d'un **signe**, sur ce qu'il **désigne**. C'est ainsi que parce que la sécheresse a été particulièrement rude sur les hauts plateaux, « un mouton », comme inscrit par un trait dans l'argile, sera cette fois une bête efflanquée, et ne pourra être rien d'autre, concrètement, dans le temps du contrat. Pour ceux de la plaine qui est restée bien arrosée, il y a divergence sur le mesurage qui vaut l'inscription de ce symbole, ou sur l'action à mettre en œuvre en regard de ce signe. On ne s'entend plus à coup sûr. L'entente sur la signification des données ne vaut qu'autant que les conventions qui la sous-tendent en arrière-plan est bonne et bien fixée et reste valable, fondée, notamment parce que les conditions du milieu ambiant sont stables ou stationnaires, ou plus simplement parce que tous peuvent les constater et s'accorder sur la signification qu'elles revêtent au regard de leurs transactions, et donc les modulations à y apporter (des corrections). L'information distante qu'apporte une tablette n'est opérante que dans un cadre fixé étroit, bien établi à l'avance et entretenu entre les parties prenantes.

Comme tout vecteur d'une langue en usage — une langue vivante —, les données sont vivantes, toute formelle que puisse paraître leur création, elles n'ont pas une signification absolue ni éternelle. Leur sens est contextualisé et donc périssable.

On demandera donc en permanence **aux hommes** aux extrémités de la chaîne d'information d'**interpréter et réinterpréter** les données, le trait dans l'argile, qui ne dit en lui-même presque rien par rapport à tout ce qu'on peut percevoir et apprécier devant la bête sur pied, et cette bête elle-même, pas dans une abstraction céleste, mais concrètement, sur une terre donnée, au bord d'un pâturage ou sur un marché, un certain jour d'une année particulière. Dans un monde changeant, une stricte assignation symbolique n'a très vite plus grand sens. « Celui-ci n'est pas pleinement un mouton, au sens implicite du contrat, tel que les données le désignent ; il est bien trop maigre. » « Allons, ami, vous savez bien qu'avec la sécheresse de cette année, vous ne pouvez pas attendre des animaux gras comme il y a deux étés. » Voilà les problèmes de mesurage (inscrire « un mouton » pour

cette bête-là), les problèmes de **contexte** dans lequel s'apprécie l'adéquation entre une donnée d'une base de donnée (ici une tablette d'argile) et la réalité concrète, la décision à prendre, l'acte à réaliser. Le déport dans l'espace de l'information, l'information retardée crée une gerbe de difficultés.

Au bilan, le code partagé qui sous-tend toute communication s'établit et reste opérant dans un contexte spécifique, et avec l'interposition de distances géographiques, mais aussi temporelles, culturelles, entre les parties prenantes, ce contexte est disloqué.

On observe également que les données ne sont pas dans la nature \cite{Kent.2000, Uexkull.1984}. Ce ne sont pas des éléments substantiels du monde. Ce ne sont pas des choses que l'on peut y saisir. L'information n'est pas non plus un élément tangible de la réalité, ou ce qu'on convient souvent d'appeler une réalité objective, elle est le fruit d'une interprétation par un individu, une symbolisation. Une chose informe en fonction de l'état de connaissance de la personne qui la déchiffre, que sa lecture s'exerce directement sur la réalité, ou sur des données (un texte, écrit dans un alphabet, un code donné, ou encore des paroles, etc.).

Les données permettent donc d'« extérioriser » le contrat, en maintenir la teneur et la mémoire hors des témoins initiaux.

Lorsque de nombreuses personnes sont appelées à intervenir d'après ce contrat, ou d'après des données qu'il a suscitées, il doit être lisible par beaucoup plus de personnes que les parties initiales, il doit devenir beaucoup plus public ; pour cela, il doit s'explicitier beaucoup plus substantiellement pour rendre accessible ce qui va de soi pour les gens de la partie.

On voit bien pourquoi le contrat n'est plus lisible, les données ne sont plus interprétables raisonnablement qu'accompagnés d'une **solide documentation** qui amène le lecteur dans le contexte de son établissement. La tarte à la crème sur ce sujet, ce sont les **métadonnées**.

Ce qu'on appelle **système d'information**, ce sont tous les éléments de cette chaîne ramifiée comprenant la mesure, l'enregistrement de données, leur utilisation pour réaliser une action. Ces éléments sont immergés dans les éléments de contexte, de pratiques de vie qui ont notamment servi de base pour s'entendre sur un code commun — ce qui pour commencer demande de dégager des objets d'intérêt, ceux pour lesquels il vaut d'établir des conventions et d'engager des transactions.

Le contexte, non décrit dans le système, permet aussi d'encadrer la communication établie par le système, et les contrats effectifs ou en puissance qu'il est susceptible de véhiculer. Il permet surtout d'ajuster les opérations de mesurage qui conduisent aux données et les actions que l'on engage sur leur foi.

On pourrait parler plus justement de systèmes de données, pour lesquels passer à l'état d'information est délicat ; il s'agit d'apporter à chaque instant et pour toutes les sollicitations qu'on leur fait subir, des éléments de mise en contexte qui permettent effectivement à l'interrogateur de construire une information adéquate sur la base des données qu'il déchiffre.

Le système d'information peut être réalisé dans des parties plus ou moins étendues sur des ordinateurs, avec des traitements tout ou partie automatisés.

Une **chaîne d'acquisition de données**, c'est, pour un utilisateur donné, toute les ramifications par lesquelles ont été constituées les données qu'il est en train de consulter, tout ce qui les détermine. Technologiquement, ce peut être une chaîne instrumentale, depuis la mesure jusqu'à l'enregistrement sur une mémoire, dans une base de données.

–2.2 . *L'histoire de l'administration est l'histoire de l'administration des données*

–2.2.1 Une organisation pour administrer des données

On s'est donc organisé avec des systèmes de données, donc de contrats, lorsqu'on s'est préoccupé du partage d'une ressource parcimonieuse, dont la disponibilité était aux yeux de tous immédiatement menacée, ou pour régler une transaction, la contrôler. Dans les deux cas, les structures administratives ont accompagné le développement de ce processus, et il s'agissait d'administrer des données. L'histoire administrative est l'histoire de l'administration de données.

L'histoire administrative a toujours procédé de cette façon : **régler un usage commun**, maintenir ou restaurer cet usage à travers la **préservation d'une ressource**. Ainsi, en Andalousie, on a instauré des tribunaux des eaux. L'administration française, de même, a mis en place une succession de dispositifs pour mesurer, enregistrer des aspects critiques de l'environnement et en arbitrer une conduite commune. Ainsi, des droits de ramassage de bois de chauffage, des droits de chasse, des hauteurs des pierres de seuil des moulins le long d'une rivière exploitée hydrauliquement, etc.

Ainsi, à une période où le gouvernement considère comme critique d'asseoir la puissance maritime de la France, et où, à l'orée de l'Empire, les surfaces de forêt sont les plus réduites qu'aura jamais connues le territoire français, et surtout les plus amenuisées en bois d'œuvre, le droit forestier, seigneurial, est placé par Colbert sous l'autorité de l'administration des Eaux et forêts. La forêt servait de nombreux usages, la fourniture du bois de chauffe, de construction, pour l'outillage, comme pâturage de réserve, domaine de cueillette et de chasse, etc. et il était défendu, sous quelque prétexte que ce soit, d'y pratiquer une coupe sans autorisation expresse de l'administration forestière ou du gouvernement : des contrats, de l'administration de données, un système d'acquisition et d'acheminement de l'information.

–2.2.2 L'état d'organisation de l'administration de données reflète les questions d'enjeu et leur histoire

À un moment donné, l'état de l'administration est donc une coupe transversale dans un processus en **évolution continue**, où des couches d'organisation successives sont sédimentées, puis remaniées, érodées, transformées, redéposées au gré de **réformes** mues par les transformations naturelles de la société qui placent des enjeux sur des aspects renouvelés, sans jamais faire table rase des structures précédentes, mais en rebâtissant au travers, à la manière dont une ville se reconstruit sur elle-même, enfouissant des couches anciennes qui attendent d'être rouvertes par l'archéologie, venant parfois y retailler pour fonder profondément de nouveaux édifices. Si à un moment donné, les structures administratives en place relèvent encore en grande partie des enjeux d'hier, c'est exceptionnellement qu'on procède à un nettoyage jusqu'à un substratum pour rebâtir de novo.

Cette histoire résulte en une **découpe administrative** intriquée en services investis de missions définies réglementairement, qui les amènent à assurer l'administration de divers jeux de données. L'état actuel de cette administration de données est le produit de la sédimentation d'années d'histoire de fonctionnement, de réformes et remaniements, de prises d'option, d'adaptations plus ou moins profondes aux mutations technologiques, considérables depuis les années 1980.

–2.2.3 La nomenclature des données : le problème de l'entrée sémantique

Les données de l'environnement sont donc aujourd'hui organisées en **thématiques** héritées de la découpe administrative historique française. La découpe administrative livre une sorte de nomenclature de fait des données de l'environnement. Elle constitue le mode premier d'entrée dans

ces données, la structuration initiale qu'elles présentent. C'est leur signification première.

Elle peut faire apparaître une multitude de services administratifs au regard de ce que l'homme de la rue considérerait comme une question unique, par exemple, une question d'assainissement (sachant justement qu'il n'emploierait probablement pas ce terme-là).

–2.3 . L'invention du développement durable, à la recherche d'une gestion synthétique

–2.3.1 D'une administration sectorielle, thématique, à une administration synthétique des données

–2.3.1.1 La prise de conscience conduisant à l'idée de développement durable : l'état critique de toutes les caractéristiques de nos milieux

On a vu comment chaque thématique d'administration (de données) a été dictée par le besoin de régler l'usage d'un aspect critique du milieu où se développait la société à ce moment-là.

Notre société est aujourd'hui marquée par la **conviction croissante** que **tous les aspects** de nos milieux, les écosystèmes qui nous hébergent sont dans cet état **critique** qui mérite, qui exige même une administration, à travers des données, pour conduire à leur égard une discipline collective, pour en faire une gestion explicite et non par des ricochets incontrôlés résultant d'actions conduites selon quelques axes de lecture isolés et indépendants.

–2.3.1.2 L'idée de développement durable pousse vers une obligation de synthèse, de gestion intégrée

La prise de conscience du développement durable consiste à comprendre qu'il faut considérer **chaque action**, de **chacun d'entre nous**, dans ses répercussions sur notre milieu à travers **toutes les échelles** spatiales emboîtées, et tous les **cheminements fonctionnels** possibles qu'elle peut emprunter, toutes les répercussions que l'on peut concevoir depuis le voisinage immédiat de l'individu jusqu'à l'échelle de la biosphère. Son point de départ médiatique date peut-être de la vue de la première image de notre planète depuis l'espace, petit îlot, arche de vie navigant dans l'immensité. Pour la première fois, on embrassait d'un seul regard toute la biosphère, mince pellicule de vie à la surface de la Terre.

La **vision cartographique** que de nombreuses facilités procurent désormais a une parenté étroite avec ce survol de la planète, elle-aussi livre une vue en rapport projectif avec la terre, qui nous paraît revêtir les attributs d'une **vue objective**, un point de vue de nulle part, ou de partout, abstrait, total. En tous cas, elle est certainement appelée à jouer un rôle central dans l'établissement d'une **vision partagée**, à la manière d'un tableau noir pour une salle de classe.

On doit se mettre en situation (ou dans la perspective) d'**évaluer** toutes les conséquences concevables de nos actions. Ce qui veut dire, en faire une **valuation**, autrement dit en donner à l'avance ce que sera leur **mesure**, et leur donner aussi une **valeur**, au sens moral du terme.

Il nous faut donc réunir l'ensemble des caractéristiques de nos milieux : toutes nos données. Plus : il faut prendre en compte toutes les **interactions**, tous les **entrecroisements** qu'elles peuvent nouer entre elles. C'est par cette **combinatoire** que passe la **vue synthétique** qu'exige le développement durable.

On part donc d'une gestion **sectorielle**, **presque par définition** de l'usage des données, qui s'exerce autour d'un ou quelques contrats circonscrits. La perspective qui s'ouvre nous entraîne dans un **entrecroisement d'échanges de données** et d'informations, à démêler, pour tenter une **gestion intégrée**.

Cela, impose de pouvoir **opérer conjointement** sur tous les jeux de données, les combiner, les composer au sens mathématique. Dans le monde des systèmes d'information, on parle d'**interopération**.

–2.3.1.3 Des contrats, des schémas de communication des données complexifiés, « totaux », en réponse à la criticité généralisée des milieux

On est donc au seuil d'une mutation, de l'administration de quelques « ressources naturelles » stratégiques, — administrer des données pour chaque ressource naturelle concernée —, vers celle d'une **administration synthétique**, synchrone, de toutes les sortes de caractéristiques de nos milieux qui nous **paraissent signifiantes**. Il faut bien noter que cette ambition de gestion est limitée dans sa portée, **relative aux caractéristiques** ou aux catégories de phénomènes **que l'on conçoit**, dans le système de pensée du moment, et non de caractéristiques qui existeraient en soi ou dans l'absolu, et qui et qui auraient le pouvoir de restituer exhaustivement la réalité. Les caractéristiques sont des concepts humains, pas des choses de la nature. Ce sont des composantes effectivement définies par un processus de mesure, qui produisent des données, concrètes, limitées dans leur portée.

Une difficulté majeure, est qu'**on ne sait pas** encore concrètement réaliser cette administration synthétique. Mais évidemment, elle passe par la mise en place de schémas de communication beaucoup plus denses et connexes que ce qu'on a su mobiliser par le passé.

Cependant l'**ordinateur** nous en semble l'outil incontournable. Ce ne sera d'ailleurs pas la première fois dans l'histoire qu'on observe que la possibilité technique vient de façon assez synchrone (peut-être l'eut-il fallu plus tôt !) avec la perception du besoin impérieux d'un outil.

Il faut aussi reconnaître que les possibilités que nous ont fait envisager la poussée technologique des ordinateurs connectés en réseaux, et leurs immenses conséquences sociales et culturelles, suscitent l'envie d'essayer de réaliser une intégration générale de données, non par nécessité, mais simplement parce qu'on en sent la possibilité à portée de main et qu'on ne peut s'empêcher de faire le test — comme l'a toujours fait la civilisation occidentale lorsque des expériences nouvelles lui semblaient envisageables. Cette possibilité active aussi le fantasme de l'**intégration** de toute la planète **en un organisme unique** (qu'on l'appelle de ses vœux ou qu'on le voue aux gémonies), épaulé par des courants sociaux comme « la mondialisation », la domination accrue de la société par des organisations, plutôt que par des individus.

Au final, on voit se dessiner la nécessité de schémas de communication entrecroisés, ou encore une **omnicontractualisation**, et une **omnipublication** de données.

C'est exactement et entièrement l'objet du schéma.

–2.3.2 Les réglementations exigeant la publication, vues comme une conséquence de ce mouvement profond

Si on observe l'avènement de textes imposant la publication des données (publiques) sur l'environnement à la lueur de ce qui précède, ils semblent s'en déduire comme une évidence.

En ce sens, la convention d'Århus, et de même, le schéma directeur des données de la Corse ne sont pas des mesures d'accompagnement d'un développement durable, mais le cœur même, la matière du processus de développement durable.

Voici un relevé des principaux textes actifs en France :

- décret \no 77-1133 du 21 septembre 1977 relatif aux installations classées pour la protection de l'environnement ;
- loi \no 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés ;
- loi \no 78-753 du 17 juillet 1978 modifiée portant diverses mesures d'amélioration des relations entre l'administration et le public et diverses dispositions d'ordre administratif, social et fiscal ;
- loi \no 79-18 du 3 janvier 1979 sur les **archives** ;
- loi \no 2000-321 du 12 avril 2000 modifiée relative aux droits des citoyens dans leurs relations avec les administrations [modifie

- notamment la loi \no 78-17 et le code pénal. « Une commission dite « **Commission d'accès aux documents administratifs** » [Cada] est chargée de veiller au respect de la liberté de l'accès aux documents administratifs et aux archives publiques. Elle établit un rapport annuel qui est rendu public. »] ;
- **convention dite d'Århus**, signée le 25 juin 1998, sur l'accès à l'information, la participation du public au processus décisionnel et l'accès à la justice en matière d'environnement <http://www.unece.org/env/pp/treatytext.htm> ;
 - directive 2003/4/CE du Parlement européen et du Conseil du 28 janvier 2003 concernant l'accès du public à l'information en matière d'environnement et abrogeant la directive 90/313/CEE du Conseil (en 1990) [exige notamment un bilan de la mise en œuvre de ses dispositions] ;
 - loi constitutionnelle \no 2005-205 du 1^{er} mars 2005 relative à la **charte de l'environnement** [la charte de l'environnement est annexée à la **constitution**, ce qui lui donne le plus haut niveau hiérarchique du droit français, supérieur aux lois notamment ; c'est un changement d'une très grande portée qui a été opéré à cette occasion, une modification historique des hiérarchies de valeurs ; qui pourra avoir de très lointaines conséquences dans la structure du droit futur] ;
 - ordonnance \no 2005-650 du 6 juin 2005, modifiant notamment la loi \no 78-753 du 17 juillet 1978 ;
 - loi \no 2005-1319 du 26 octobre 2005 portant diverses dispositions d'adaptation du droit communautaire en matière d'environnement [dont sont issus les articles L. 124-1 à L. 124-8 du code de l'environnement] ;
 - décret \no 2005-1755 du 30 décembre 2005 relatif à la liberté d'accès aux documents administratifs et à la réutilisation des informations publiques, pris pour application de la loi \no 78-753 du 17 juillet 1978 [!] [Prévoit qu'un grand nombre d'autorités publiques désignent une **personne responsable de l'accès aux documents administratifs** et des questions relatives à la réutilisation des informations [sic] publiques ; dans tous les cas, une autorité publique doit désigner officiellement une **personne responsable de l'accès à l'information en matière d'environnement.**] ;
 - décret \no 2006-578 du 22 mai 2006 relatif à l'information et à la participation du public en matière d'environnement, modifiant le code de l'environnement [articles R. 124-1 à R. 124-5]
 - loi \no 2006-686 du 13 juin 2006 relative à la transparence et à la sécurité en matière nucléaire [précise le droit à l'information en matière de sûreté nucléaire et de radio-protection] ;
 - [code de l'environnement, (livre I^{er}, titre II, chapitre IV) articles L. 124-1 à L. 124-8 et R. 124-1 à R. 124-5, prévoit certaines modalités imposées par la convention d'Århus et le droit communautaire ;]
 - circulaire du 18 octobre 2007 relative à la mise en œuvre des dispositions régissant le droit d'accès à l'information relative à l'environnement, NOR : DEVGo700268C (ministère de l'Écologie, du développement et de l'aménagement durables, non parue au Journal officiel).

Les textes concernant l'environnement sont désormais codifiés, pour une version lisible de plusieurs lois et décrets précédents, il est préférable de se reporter à un code à jour. Les versions officielles, initiale et consolidée des lois, décrets, etc. sont accessibles sur <http://www.legifrance.gouv.fr>.

Il faudra en donner un commentaire informatif, qui ne prétend pas se substituer au texte, tout en s'efforçant de lui être aussi fidèle que possible.

-2.3.3 D'un milieu, une nature adverse qu'il s'agit de rendre à nos raisons à une nature avec laquelle on converse

Souvent, notre milieu a été perçu, conçu comme un espace extérieur, à affronter, à modeler, qui bien que pouvant être aménagé selon notre volonté, se comportait essentiellement comme s'il était infini devant nos

actions : on a considéré les transformations qu'on y exerçait étaient tamponnées localement ; ou encore, que les répercussions des aménagements qu'on lui imprimerait se cantonneraient aux limites idéales d'un confort et d'une sécurité accrue tels qu'on les avait rêvés, aucune autre répercussion

L'importance qu'a prise l'espèce humaine, le déploiement de ses actions permettent de constater que désormais, chaque geste de chacun d'entre nous a des résonances globales. Ce n'est pas seulement l'idée de principe que toute la biosphère forme un seul système en interaction, c'est le fait palpable de changements de grande ampleur qui remettent en cause notre confort et notre sécurité, et dont on attribue la responsabilité à l'activité humaine.

L'idée du développement durable procède de la reconnaissance qu'avec l'importance de la population humaine, chaque action, si petite soit-elle, engendre des conséquences significatives de toutes sortes, parce qu'elle en côtoie des millions d'autres.

L'action humaine et les processus de la biosphère planétaires devenus commensurables, ce milieu de vie qu'on a vécu comme extérieur change de statut. Il n'est plus distinct mais **conjoint**. Il n'est pas étranger, on est avec. On n'est plus en affrontement, ou retranché, ou coupé de lui, mais en interaction. Il n'est pas adverse, on est ensemble sur le même fil d'équilibriste.

On a appris à concevoir et opérer sur des détails séparés, il faut maintenant penser chaque geste dans la globalité. Il faut apprendre la vision synthétique.

Finalement, il était très candide d'espérer qu'on pouvait aménager l'espace, agir violemment dessus, et que la nature ferait toujours pour le mieux pour nous, se rendant à notre volonté sur les paramètres de nos desiderata, sans rien altérer de tous les autres aspects de la niche écologique qu'elle nous offre. On a voulu se convaincre que les risques naturels s'estomperait devant notre désir, plieraient à notre volonté, que finalement la rationalité de notre confort conduirait la marche des choses de la nature.

Dans cette interaction, cette co-détermination permanente, il est impropre de parler de préserver son environnement, car cela voudrait dire pouvoir se retirer ailleurs, se soustraire du paysage. C'est par défaut comme cela que l'on a considéré jusque maintenant notre rapport avec la nature. Il faut vivre avec, compatiblement, et poursuivre une relation viable de détermination mutuelle.

À proprement parler, « gérer » un milieu, ce n'est pas l'exploiter, c'est y vivre, établir un régime d'équilibre, de **co-détermination** avec lui. Il semble qu'aujourd'hui, cela veuille dire **aussi**, essentiellement, faire vivre sa représentation symbolique, sur des mémoires externes : **administrer des données**, et à travers elles, administrer autant d'aspects de nos milieux.

-2.3.4 Les défis de la gestion synthétique des données, de l'administration synthétique

-2.3.4.1 Le développement durable passe la réalisation de ces schémas : c'est donc avant tout un problème de communication

Construire des vues synthétiques suppose un **accès fluide aux données**, réalisé par ces schémas de communication.

La fluidité, c'est la **vitesse** à laquelle on accède aux données, une propriété essentielle de la relation qu'on peut entretenir avec elles. Contrairement à la première intuition, la vitesse n'est pas un paramètre quantitatif dans ce contexte, c'est un paramètre qualitatif, c'est-à-dire que des variations de vitesse changent radicalement la nature de la relation avec les données. Selon la situation, quelques secondes ou quelques minutes de délai peuvent suffire à signifier en pratique « données inaccessibles donc inutilisables, abandon », parce qu'il est impossible à l'utilisateur d'attendre autant pour

pouvoir prendre connaissance des données, et se construire une information.

C'est aussi ce qui fait que le web change radicalement la donne en matière d'information, même à contenu de données inchangé. Imaginer qu'il existe dans une bibliothèque de telle ville d'Europe, un ouvrage sur le sujet qui nous préoccupe est une chose, savoir qu'il y existe effectivement, pouvoir consulter le catalogue, pouvoir prendre connaissance de l'ouvrage, sont des niveaux qualitativement très différents. Selon qu'il faut entreprendre un voyage long et coûteux simplement pour savoir si..., ou que c'est possible en quelques échanges de courrier, ou encore en quelques clics sur un ordinateur — et cela d'un geste pour une myriade de bibliothèques du continent —, on passe dans presque tous les cas de la velléité (zéro information) à la lecture effective. Même si dans tous les cas, les données disponibles sont strictement les mêmes.

À un certain degré, lenteur, viscosité dans un canal de communication équivaut à **opacité**. Publier sans offrir de rapidité, ou publier selon des modalités qui interdisent la manipulation intégrée des données, ce n'est pas publier.

Il faut donc que les données soient **matériellement accessibles**, il faut qu'elles soient **intellectuellement accessibles**.

L'accès matériel doit être fluide et uniforme, sous une forme qui permet des **traitements automatiques**. Cela conduit, sans exclure les modalités classiques d'accès, à les procurer sur le web.

Le critère du web ne suffit pas, il faut pousser à fond ses principes d'intégration en couches d'interaction car là-encore, diverses manipulations de rapatriement pour utiliser les données, ou accès dynamique comme le permettent des services \wfs\ et \wms\ placent l'utilisateur dans deux mondes d'information complètement différents.

Ce sont les **couches de données** du système (choix d'encodage, formats de fichiers, couches de réseaux, coiffées par les couches supérieures du web, et les normalisations supplémentaires comme celles du consortium ouvert de l'information géographique, l'\ogc\ — Open geospatial consortium) qui permettent cette accessibilité matérielle.

Cependant, l'exigence va plus loin. On a vu que le **système de données** se complète en **système d'information**, lorsqu'un individu interprète les données : chacune de ces prises de connaissance s'exerce comme une **couche d'information** du système, qui coiffe toutes les couches de données. Il faut donc que les données soient aussi **intellectuellement accessibles**, c'est-à-dire que l'on puisse d'une part comprendre le rapport qu'entretient un lot de données avec la réalité que l'on cherche à appréhender dans une situation donnée, et d'autre part que l'on puisse, face à une question formulée de façon arbitraire, découvrir aisément quels éléments sont disponibles pour l'étayer.

On peut distinguer schématiquement la **fiche d'identité** des données, et le **contenu même** des données. La fiche livre une partie de ce qui est indispensable pour pouvoir interpréter ce qui peut n'être sinon qu'une suite ininterrompue de nombres dans le cas d'un fichier de données. C'est aussi vrai dans le cas d'un fichier de texte, dont la lecture peut prendre un tout autre sens selon la façon dont on l'insère dans un contexte général.

Pour se fixer les idées, dans un courrier, la fiche d'identité peut correspondre aux informations situées dans l'en-tête : qui s'adresse à qui, à quelle date, l'objet, les références, etc. Le reste est le contenu du courrier au sens strict, un message. En réalité, le courrier n'existe comme message que s'il est bien muni de la totalité de ces attributs.

Beaucoup d'autres éléments manquent pour qu'un lecteur soit en mesure d'apprécier la teneur du message ; au minimum, tout ce qu'on trouverait typiquement dans le dossier où il serait classé : les échanges précédents, des notes, des commentaires qui le situent dans le déroulement d'une relation entre les parties.

Dans le monde des systèmes d'information, on appelle **métadonnées** toutes les **données formalisées** correspondant à la fiche d'identité d'un message. Elles sont formalisées, parce qu'il y a un **accord assez général** sur leur **codification** (par exemple expéditeur, signataire, date, objet, etc. pour un courrier ; pour une donnée cartographique, les coordonnées, les dates d'édition, etc.).

Les métadonnées sont donc la frange qui peut être codifiée des indications de contexte indispensables pour identifier ce qu'est une donnée, et envisager de l'utiliser, de l'interpréter. Elles font partie de l'information qui peut être attachée directement à une donnée (elles en sont normalement partie intégrante, comme dans le cas du courrier, elles en sont indissociables pour quelque usage qu'on voudrait en faire). De nombreux efforts de coordination, de normalisation, se poursuivent pour établir des conventions efficaces sur la dénomination de métadonnées pour permettre de rechercher **automatiquement** des données présentant telle et telle caractéristiques en interrogeant des systèmes d'information (une de ces initiatives est le Dublin core). Ces conventions, ces normes forment dans les systèmes de communication des couches supplémentaires de **conventions de langage** qui sont de même type et complètent celles des **couches de données**.

Si l'on pousse à fond la définition couramment proposée du terme métadonnée, une « **donnée sur les données** », qui correspond à sa construction étymologique, et qu'on y met tout ce qui peut être utile à l'interprétation d'un message particulier, alors, comme on l'a illustré dans le cas d'un courrier, tout le dossier qui va avec est métadonnée, tout le corpus de données sous une forme ou une autre que l'on peut rattacher à celle que l'on a en mains joue pour elle le rôle de métadonnée : il n'y a **pas de distinction de nature** entre des données et des métadonnées, les métadonnées sont avant tout des données, et, suivant le point de vue de lecture, le même fragment peut être l'un ou l'autre.

La sphère des métadonnées qui peut être codifiée correspond aux structures d'échanges communes, répétitives, donc convenues ; elle se résout en une nébuleuse qui peut, dans une situation particulière de transaction, pousser des ramifications absolument partout.

Les métadonnées sont donc **avant tout des données**, et, comme telles, répondent aux mêmes conditions d'établissement, de qualification pour un usage.

L'**informatique**, soit le **traitement automatique de l'information**, serait plus justement nommée donnée-tique, à l'instar du terme danois *datalogi*, ou encore de la *databelhandling* des norvégiens à l'université d'Oslo, — traitement, manipulation de données. La confusion entre donnée et information date peut-être du glissement de terme à partir de l'ouvrage de Shannon et Weaver *The mathematical theory of communication*, dont est issue la discipline qu'on appelle désormais théorie de l'information.

Le traitement automatique de données est possible — par exemple entre deux couches de données qui s'interfaçent ou communiquent --- pour peu que la **matériau** qu'elles échangent, c'est-à-dire du **texte**, se contraigne à une **syntaxe stricte**, par rapport à celle des langues naturelles. Cette restriction fait perdre en capacité d'expression, mais en contrepartie on peut se décharger sur une machine de traitements dont **l'issue est garantie**, sur lesquels on peut donc s'appuyer avec **certitude**.

C'est pour cela que les **langages informatiques** sont affaires de **codages, formats** (de fichiers notamment), **protocoles, communications** (car un langage implique toujours une forme de communication). Et réciproquement, un format de fichier est une sorte de langage, une convention sur la base de laquelle on peut échanger des données. Avec à chaque fois, la nécessité d'une entente entre parties.

Pour le schéma, cela veut dire que les nomenclatures, en ce qu'elles sont passibles de généralité, doivent se conformer à celles adoptées par ailleurs,

et en ce qu'elles doivent être spécifiques à la Corse, elles doivent émaner d'un accord entre les partenaires du schéma.

L'ultime mise en contexte, pour interpréter une donnée ne peut relever de langages formalisés, elle sollicite les facultés d'un individu pour ré-immérer la perception de la donnée dans le monde (qui n'est pas formel : il ne respecte pas une syntaxe de langage informatique).

C'est le moment où l'on a épuisé ce que pouvaient faire des manipulations automatiques, grossièrement, présenter fiablement des données codifiées. Il reste alors à les interpréter, et cette opération ne peut pas ressortir à un système automatisé. Elle ne peut être le fait que d'un individu. Corrélativement, il peut avoir à mobiliser d'autres **informations** que ce que peut lui procurer un jeu de métadonnées codifié. Il dépasse en cela ce qu'exploite un système automatique.

-2.3.4.2 La difficulté des schémas de communication ouverts à tous : la pluralité des points de vue

Rendre les données intellectuellement accessibles se heurte à un autre obstacle, celui de la pluralité de points de vue.

Notre système de données s'achève toujours en système d'information par l'interprétation faite par un individu.

Mais chaque individu est lui-même son propre **référentiel d'interprétation**, et il n'est pas facile d'aviser quelqu'un de façon à être certain qu'il produira une interprétation conforme aux concepts de celui qui émet le message, qui produit les données.

Or dans le projet du développement durable, il s'agit bien de permettre à **tous d'accéder à l'information**, au sens de la rendre **compréhensible pour tous**, car s'il impose une synthèse pour fonder toutes les décisions, alors tous doivent pouvoir accéder à la synthèse, et la vérifier pour leur propre compte, voire la réaliser eux-mêmes.

Pour permettre cela, il faut que l'on ait réussi à réduire les différences de langage, le **syndrome de Babel**. Un pêcheur n'a pas la même rivière qu'un hydraulicien, ni qu'un botaniste, ni qu'un agriculteur, ni qu'un baigneur. Parlent-ils seulement de la même chose ?

La condition de réussite du projet de développement durable n'est pas dans une innovation technique encore à venir. L'enjeu clef est la capacité à communiquer effectivement entre personnes, c'est-à-dire à disposer effectivement d'un **langage commun**, le seul moyen de faire se rencontrer une **pluralité de points de vue**, le seul moyen de construire une synthèse avisée, d'envisager une complexité ouverte.

Au final, le problème clef de l'avènement d'un développement durable est un problème **pédagogique**, voire **encyclopédique**, puisqu'il faut parvenir à **fonder une base commune** de connaissances.

Le défi majeur du schéma directeur des données de l'environnement de la Corse est donc de parvenir à rendre **effectivement accessible aux personnes**, sur les couches supérieures du système d'information, **tout ce qui est perçu** sur l'environnement.

Ce peut être **perçu par des personnes** en chair, qui doivent pouvoir abonder le débat de leur perception qu'ils soient pêcheur, hydraulicien, botaniste, agriculteur, baigneur, ce peut être **perçu par des systèmes organisés (mesuré)**, comme ceux que mettent en œuvre les acteurs premiers du schéma directeur.

-2.3.4.3 La pluralité des points de vue et la déconnexion des connaissances

Le même déterminisme qui en réponse à l'exploitation d'une ressource rare, a produit tant de fois une administration de données est donc à l'aube de provoquer le déploiement d'une administration intégrée de toutes les données sur toutes les ressources. Cela se produit alors même que la connaissance humaine n'a probablement jamais été si déconnectée, parcellisée \cite{Davis.Hersh.}.

Très peu de personnes aujourd'hui connaissent très bien leur milieu de vie, ne serait-ce qu'aussi bien que la génération précédente. En fait leur milieu de vie n'est plus du tout ce qu'un habitant du « même » lieu au début du 20^e siècle aurait pris pour objectif, incontournable tant il est matériel : vallée, rivière, champs, chemins, arbres, parfums. On n'habite plus un endroit, on vient seulement y dormir.

On peut même dire que beaucoup de personnes, au contraire, connaissent très mal leur milieu, puisqu'elles arrivent sans s'en douter à installer leur maison dans le lit d'une rivière. Chaque inondation en donne des exemples à profusion.

On est à des lieues de la structure de connaissance comme elle existe dans ce que rapportent les anthropologues des peuples de chasseurs-cueilleurs contemporains, en Papouasie-Nouvelle Guinée par exemple, qui nous paraissent vivre dans **un monde très unifié**, très connecté, où chacun de leurs gestes s'inscrit dans ensemble synthétique cohérent, et où la base de savoirs est partagée par la plupart, comme en témoigne Jared Diamond parmi d'autres dans son *Guns, germs and steel*. Assurément, ils sont ainsi plus proches de la réalisation d'un **mode de vie durable** que ce que nous procure notre **développement**.

Nos sociétés ont pu arriver à un stade d'édification où il est devenu possible d'accommoder une myriade de niches écologiques où chacun peut se **spécialiser** à loisir. Du coup plus personne ne sait ce que sait un autre, et sûrement pas son plus proche voisin. On peut se permettre des talents si spécifiques, si étroits, des incompétences si notoires, des lacunes, des faiblesses si immenses que jamais nous n'aurions survécu de cette façon en d'autres temps. Nous sommes donc dans une époque de désintégration des savoirs, au sens étymologique, ils sont déconnectés. Jamais les espaces de savoir n'ont été aussi immenses, mais il n'y a pas assez d'humains pour tous les occuper, et deux personnes, tirées au hasard dans cette population sont dans l'impossibilité d'échanger des propos sur la quasi-totalité de leurs savoirs respectifs \cite{Davis.Hersh}. C'est bien le stade avancé de l'édification de la tour de Babel.

On se rêve dans un immense monde unitaire, interconnecté, vaste organisme vibrant à l'unisson, en même temps, pour l'heure, chacun voit midi à sa porte : chacun est focalisé sur des caractéristiques qui le motivent particulièrement. Comment vaincre la muraille du langage ? Le défi de communication que pose le développement durable n'est pas petit.

Le pêcheur veut une rivière à poissons, l'hydraulicien voit ses biefs, le botaniste s'inquiète de la ripisylve et des plantes aquatiques. Qui sait comprendre les signes que l'autre voit comme des évidences, brandis comme des sémaphores dans le milieu ? Pour comprendre les signes de l'autre, pour comprendre ses mots, il faut venir faire ses gestes, vivre un peu sa vie.

W. Sombart, De l'homme, cité par Uexküll \cite[p. 97]{Uexkull:1984}, l'exprime très clairement : « Il n'existe pas de forêt en tant que milieu objectivement déterminé il y a une forêt-pour-le-forestier, une forêt-pour-le-chasseur, une forêt-pour-le-botaniste, une forêt-pour-le-promeneur, une forêt-pour-l'ami-de-la-nature, une forêt-pour-celui-qui ramasse-du-bois ou celui-qui-cueille-des-baies, une forêt de légende où se perd le petit Poucet. »

Tous ces mondes sont différents, et bien que s'interpénétrant, essentiellement étrangers les uns aux autres (et nous n'avons parlé que des mondes humains ! que dire des mondes des animaux, des plantes, des bactéries, des champignons ?). Le défi du développement durable est d'établir des langages pour les faire converser.

-2.4 . En résumé (vocabulaire)

Toute **communication** passe par l'échange de **données** selon un **code**, et suppose une sorte de contrat, en même temps qu'elle supporte l'exécution de contrats, c'est-à-dire d'actions coordonnées, concertées entre au moins deux parties.

Il y a donc au moins un contrat sous-jacent à toute administration de données, toute administration. Pour comprendre à quel usage précis peuvent servir les données, il importe d'exhumer la teneur exacte de ce contrat, et de tracer son évolution au cours du temps, qui a pu, de transmutation en réforme, en modifier radicalement la donne.

On appelle **donnée** toute séquence de signes dans un encodage déterminé selon un code, une convention établie entre des parties prenantes, qui l'ont donc adopté comme support de communication, pour s'envoyer des messages, communiquer. Elles se communiquent des **données** qui leur apportent, lorsqu'elles les interprètent, des **informations**. Un jeu de données est donc nécessairement, avant tout, un **flux de données**.

Noter que ces définitions placent un peu en porte-à-faux l'expression « système d'information », qui serait plus proprement qualifié de **système de données**.

Cette présentation d'ensemble met en place les éléments utiles pour raisonner sur un schéma de données, et livre presque directement des questions clefs à se poser concrètement dans la démarche de sa mise en place pratique. Le développement qui suit apporte des compléments, notamment sur le point crucial de l'usage des données.

–3 Les principes auxquels on ne peut déroger pour publier des données

Ici il s'agit de reprendre les principes tels qu'évoqués dans la note de cadrage du schéma. Quelques paragraphes les relient à la mise en perspective qui en est faite dans la section –2.

–3.1 . Publier largement, rendre disponible à tous, c'est créer un vaste système de communications

Un usage approprié de données sur l'environnement pour engager chaque action de façon pondérée, nous place devant la perspective d'un entrecroisement vertigineux d'échanges de données, de communications, dont peut résulter, simplement par inflation de leur nombre, une confusion digne de la tour de Babel.

Jusqu'à présent, on a réglé des communications en abstrayant des éléments de la réalité perçue sous forme de codes.

On en est aujourd'hui à un stade où ces dispositifs d'administration de données eux-mêmes doivent connaître une abstraction, une explicitation, un typage, une formalisation, la mise dans une perspective plus vaste et intégrée, et une organisation en couches. En fait il faut une administration de ces administrations de données. Il faut les faire entrer toutes dans un schéma cohérent, leur donner une cohérence globale.

Créer des communications de communications, c'est-à-dire mettre en place une, des couches supplémentaires par lesquelles les systèmes de données en usage actuellement deviennent capables d'entrer eux-mêmes.

C'est une sorte de couche supplémentaire que l'on confère à l'administration des données.

–3.2 . Pour organiser des structures de communication complexes, adopter une disposition en couches

C'est une architecture qui semble avoir été explicitée pour la première fois en informatique par Edsger Dijkstra en 1968, démontrée dans son système d'exploitation « THE », mais plus immédiatement encore, c'est aussi comme cela que les biologistes conçoivent et décrivent les niveaux d'organisation du monde vivant.

Elle a beaucoup contribué à structurer la pensée, en donnant un cadre pour se représenter la construction de grands programmes et systèmes informatiques, en formulant une organisation qu'on arrive à concevoir et à maîtriser, un point de mire à viser. C'est dans le domaine des réseaux informatiques qu'elle a connu le succès le plus évident, puisque ceux-ci

sont explicitement construits en couches empilées. La couche supérieure est connue (superficiellement) de tous, c'est celle du web (le protocole http ou ses homologues).

L'organisation en couches apporte une propriété très puissante dans un système de communication : l'abstraction. Une des caractéristiques de l'abstraction est l'oubli. Deux choses peuvent s'observer comme étant différentes, mais à un certain niveau d'abstraction, relever de la même catégorie. Le niveau d'abstraction **oublie** les détails qui les rendent distinctes.

Par un mécanisme tout à fait analogue, une couche d'un système de communication converse avec les couches homologues d'un autre système en faisant abstraction de l'intervention de toutes les couches sous-jacentes. On dit souvent qu'elles sont **transparentes**.

Pour ces raisons, étayées au §1, le schéma des données de l'environnement de la Corse ne peut déroger à une architecture en couches, fondée sur le web, puisque c'est la couche qui coiffe les protocoles des réseaux de communication d'usage universel, existante, disponible. Plus précisément, le schéma se fonde sur des **services informationnels** établis sur le web. En particulier, la **présentation des services d'information géographique** s'appuie sur les normes \wms\ et \wfs.

-3.3 . Les raisons de migrer, et de ne pas migrer vers la configuration de principe (non intrusion, non destruction)

Dans les principes généraux qui ont été décrits,, les notions énoncées d'**unicité** de la donnée, de **répartition**, de **subsidiarité**, d'**interopérabilité**, qui se recouvrent entre elles en certains points, trouvent une place logique.

L'adoption du schéma touche et questionne à différents niveaux les systèmes d'information de chacune des parties prenantes : infrastructure technique et humaine, organisation, mais aussi missions et positionnement.

Elle impose aussi des changements dans la réalisation interne du flux de données, idéalement aussi discrets que possibles, intéressant seulement un interfaçage permettant de satisfaire les modalités de présentation selon des services web, et les normes de documentation du schéma.

Selon les données (qui, en fonction de l'implémentation concrète, peuvent être vues par les utilisateurs plus comme une application), cela peut impliquer une réforme plus ou moins radicale ; dans l'exemple de \gsp, présenté un peu plus loin, cela supposerait une réécriture complète du logiciel pour une intégration totale. Probablement, si on en venait au moment de publier les données correspondantes, on serait tenté de se cantonner à l'écriture d'une interface automatisant l'export des tables de bases de données (et de documentations) dans un format de présentation convenable pour une exploration au travers du web.

Ainsi, la capacité d'un jeu, d'un flux de données à être amené au schéma, c'est-à-dire la facilité de l'adaptation ou de la conversion nécessaire pour le faire, doit-elle être confrontée à l'exigence et l'urgence de publicité des données, combien sont partagés avec des partenaires directs ou sa communication auprès du public est cruciale. Chemin faisant, des questionnements sur la réorganisation profonde de ces flux, pouvant même mettre en cause la pertinence de leur maintien peuvent voir le jour.

Le travail d'inventaire, permanent, s'assortit du soupèsément de l'**opportunité** des adaptations à pratiquer pour satisfaire au schéma : tel flux de données, et le système de communication, le contrat — les contrats — qui lui correspondent, sont-ils toujours d'actualité sous cette forme ? C'est une idée pour laquelle le jargon des systèmes d'information emploie souvent l'expression de **principe de proportionnalité**, à satisfaire, entre les efforts à entreprendre et les objectifs qu'ils permettront de remplir.

Dans le cas limite, seule une url assure ce transfert, toute l'infrastructure technique reste entièrement à la charge du partenaire dépositaire de la production d'un jeu de données. Le schéma se réalise alors exclusivement dans sa forme idéale, c'est-à-dire comme des conventions sur des modalités de communication entre l'ensemble des partenaires.

Si des données sont accessibles à la fois sur le système propre d'un partenaire et celui de l'infrastructure commune, il importe qu'elles y apparaissent sous la forme d'un identifiant unique (typiquement une url commune), sinon cette multiplication de sources amène l'utilisateur à des vérifications fastidieuses pour s'assurer de leur identité.

L'adoption de ces principes peut être très bénéfique pour chaque système d'information sectoriel, propre à un acteur.

Pour un certain nombre de jeux de données, l'administration au travers d'un service web d'information est un avantage immédiat, surtout en phase de réorganisation de services administratifs (adopter des principes contraires a été une façon de stabiliser, cristalliser des services administratifs). L'organisation en services web fait apparaître une fluidité là où peut se situer habituellement beaucoup d'inertie.

-3.4 . La qualification des données

-3.4.1 L'administration de données, comparée à une boucle perception-action

Il faut voir la gestion des données comme une boucle perception-action, seule l'action est une confirmation d'une validité quelconque de ce qui a été perçu. Il n'y a pas de vérité ou de validité en soi du perçu, ou des données, il y a leur appropriation éprouvée par l'expérience des faits à un certain usage.

Les données collectées dans la perception sont des hypothèses. Tant qu'on n'engage aucune action, rien ne permet de confirmer ni d'infirmer ces hypothèses. L'action est lancée comme un pari sur la base de ces hypothèses, et l'issue de l'action dit en retour si les données perceptives permettaient ce pari, et ce pari seulement. Elles ont prouvé qu'elles étaient propre à cet usage. C'est la seule façon de certifier une perception, c'est la seule façon de certifier une donnée.

Autrement dit, si on n'utilise pas des données, on n'a aucun retour, on ne peut pas les garantir. On ne peut certifier une donnée qui n'est pas utilisée.

Dans une base de données, on observe, naturellement, que tout ce qui ne sert pas tend à être abandonné ou négligé, et ne peut être garanti.

Cependant, une inspection même rapide, par quelqu'un qui connaît bien son métier, doit être considérée comme une utilisation virtuelle. Implicitement, il se projette dans un usage, et il dit « je n'utiliserais pas ça », ou « ça a l'air d'aller » ou encore « ici, il y a une valeur aberrante ».

Une donnée n'est en soi ni juste, ni erronée, elle possède un certain domaine de validité, elle est appropriée à une certaine palette d'usages.

Une donnée, numérique ou symbolique a une apparence trompeuse. Par exemple, une température relevée en Celsius, mettons 40, a tous les dehors d'un nombre comme un autre. Pourtant, si $40=2 \times 20$, on ne peut légalement dire que 40\degres font deux fois 20\degres : la représentation de la donnée ne suffit pas à dire ce qu'il est valide de faire avec. Couplé à une donnée, il faut procurer un contexte qui explicite son domaine de validité, dans notre exemple des températures, donner notamment les opérations mathématiques autorisées (les invariants de l'échelle). Cet inventaire de propriétés est ouvert, on ne peut venir à bout de caractériser ce qu'on peut faire avec des données, ni en quoi on peut compter dessus, et chaque jour amène de nouvelles idées de manipulations ou d'applications.

La réglementation impose la transparence, la publicité des données de l'environnement. Mais la transparence bien comprise suppose que l'on accède à la signification des données. La caractérisation ultime de la

signification d'une donnée est ce qu'il est valide de faire avec : ce que l'on peut décider en l'utilisant, quels traitements il est licite de pratiquer dessus, toujours ultimement en référence à un usage qui permet de vérifier sa valeur de prédiction.

La seule **stratégie durable pour caractériser des données** consiste à **documenter** aussi précisément que possible leur **chaîne d'acquisition** : le dispositif de mesure, et toute la chaîne de transformations qu'elles empruntent, y compris les motivations et les justifications des choix qui ont conduit à cette mise en place.

Les métadonnées, au sens employé aujourd'hui, sont une façon de structurer une partie cette information, pour ce qu'elle a de formalisable, ce qui permet des requêtes automatiques. Les métadonnées servent à savoir ce que sont les données, autrement dit à réduire les différences de langage. Mais les métadonnées sont elles-mêmes des données. Si elles ne servent pas à leur gestionnaire, elles ne pourront être garanties, elles ne seront pas renseignées.

Le pari de la perception correspond au contrat entre le producteur et l'utilisateur de données. Mais ce contrat ne peut durer qui si on entretient une boucle de rétro-action qui confirme à tout moment la validité des données par l'usage qui en est fait. C'est la notion essentielle de boucle d'asservissement, découverte par la cybernétique.

La rétro-action est beaucoup plus difficile à mettre en œuvre quand les différents éléments de la production et de l'utilisation des données sont égrenés le long d'une chaîne. Il faut, justement, écrire entre chaque couche des contrats qui règlent leurs échanges, et réaliser effectivement des passages d'information conformes à ces contrats. En revanche, dynamiser l'utilisation des données en les rendant largement disponible, c'est bien sûr devoir faire face à de nombreux retours d'information, mais en contrepartie chacun d'eux est une forme de certification des données.

-3.4.2 Écoutons à ce sujet Laurent Bloch, dans son remarquable ouvrage dont la lecture devrait être interdite

\cite[p. 134–135]{Bloch.2005}

(Avec l'espoir que la formulation du titre précédent est assez provocante pour vous vous décider enfin à suivre les suggestions de lecture qu'on s'efforce de vous faire :-)

Un exemple, dont le sujet n'est pas complètement étranger à ce qu'ont à faire les acteurs primaires du schéma...

« Dans plusieurs organismes successifs, j'ai eu la responsabilité des systèmes et des réseaux informatiques, et j'ai été confronté à la nécessité d'avoir une idée la plus précise possible du parc de micro-ordinateurs, imprimantes et autre réseaux locaux installés par les utilisateurs, afin d'évaluer correctement la dimension des infrastructures à installer. Le problème semblait minuscule et simple, et les gestionnaires du lieu m'ont conseillé d'explorer pour ce faire le fichier des immobilisations où étaient enregistrés tous les matériels acquis par l'organisme et suffisamment onéreux pour avoir le statut d'investissement.

La méthode ainsi suggérée soulève dès l'abord une objection de fond : un inventaire des immobilisations omet, par définition, tous les matériels loués ou achetés en crédit-bail, que la comptabilité connaît sous la rubrique « Travaux, fournitures et services extérieurs », très éloignée des investissements, de telle sorte que le rapprochement et la mise en cohérence de ces deux sources est pratiquement impossible. Sans aller plus loin, on voit déjà que le responsable informatique, à la recherche de données relatives aux objets physiques de son domaine, trouvera difficilement ce qu'il cherche dans les informations de gestion. »

Il poursuit :

« L'enregistrement des immobilisations est une obligation légale, mais les données ainsi collectées sont rarement exploitées. Des données

inexploitées ne stimulent pas la mise au point de bonnes procédures de collecte : partout où j'ai eu l'occasion de les examiner, les données relatives aux immobilisations étaient de très mauvaise qualité. Les montants étaient à peu près exacts, mais les intitulés censés décrire les objets acquis étaient totalement inexploitable, rédigés par des personnes qui n'avaient pas la moindre idée de ce dont il s'agissait ni de comment [sic] se rapprocher d'une nomenclature systématique. De toutes les façons, les utilisateurs s'étaient ingéniés à acheter des matériels parfois importants en mettant à contribution les sources de financement les plus variées, dont certaines n'entraînaient aucune inscription dans les livres de l'organisme. Les matériels acquis avec l'argent de contrats de recherche divers et variés n'étaient même pas la propriété de l'entreprise, mais il me fallait néanmoins les raccorder au réseau, et donc les connaître. Bref, l'exploitation des bases de données opérationnelles ne donnait aucun espoir d'obtenir une idée même approximative du parc installé, alors que le nom des données en laissait espérer une description exacte. »

Ces quelques lignes soulèvent à peu près toute la gamme des problèmes de l'utilisation de données, ou de ses propriétés caractéristiques :

- une donnée ne vaut que ce à quoi elle sert, c'est cela qui la qualifie ;
- d'où la difficulté d'utiliser une donnée à autre chose que ce à quoi elle sert ;
- et celle de se fier à l'identification d'une donnée sur la seule base de son nom.

Un peu plus loin (p. 137), il rapporte la construction d'un questionnaire puis l'organisation de l'enquête annuelle d'entreprise correspondante.

« De cette expérience de construction d'un système d'information — en effet qu'est-ce qu'une enquête statistique, sinon un SI ? — nous avons retenu que pour obtenir l'accès aux données de base, et pour en extraire l'information que nous voulions, il nous a fallu d'une part nous livrer à une étude très spécifique de l'univers particulier que nous nous proposons d'explorer, d'autre part mener des négociations somme toutes agréables bien qu'assez serrées avec les membres de cet univers.

Lorsque les instituts de statistique publics traversent des saisons d'étiage budgétaire, ils n'ont plus les moyens d'organiser toutes les enquêtes qu'ils souhaiteraient, et ils se rabattent sur l'exploitation des données administratives. Parfois cela donne des résultats exploitables, mais si cette pratique est peu prisée c'est bien parce que la plupart du temps les données extraites des bases administratives ne satisfont pas aux impératifs de la statistique. »

—4 Faire évoluer l'administration des données vers une configuration qui réponde à ces principes

Le processus du schéma est conduit en 2008 de telle manière que le pronostic de sa réussite est très favorable. Des infrastructures techniques répondant aux principes énoncés sont en places, une batterie respectable de données y est disponible, notamment les principaux fonds cartographiques, indispensables à tous pour mettre en scène leurs données géographiques. On peut facilement faire des démonstrations de réalisations d'applications d'édition de bases de données à travers une interface web de système d'information géographique. On peut appeler les mêmes services de données depuis ses applications de \sig\ habituelles. Le dispositif en place permet donc à tous de percevoir concrètement ce qui est rendu possible par l'infrastructure web du schéma, et partant, la démarche consistant pour chacun à faire le point de son propre système pour étudier ses possibilités d'adaptation à ce nouveau type d'architecture peut se mettre en route.

Nous avons réuni ici quelques remarques, toujours rédigées dans la continuité de la section –2, qui pourront le cas échéant être utiles dans la suite.

Le schéma vise à énumérer, documenter, faire migrer, et donner les moyens techniques d'une souplesse administrative, et d'une intégration.

–4.1 . Vers un inventaire permanent des composantes entrant dans le schéma

Il y aura, il doit y avoir un inventaire permanent des composantes du schéma. Tout d'abord, pour avancer dans l'immense stock de données et procédures sédimentées aux cours d'années d'évolution. Puis pour accompagner les reconfigurations que l'on fera subir aux chaînes de gestion de données, et les évolutions de sémantique qui se feront à mesure que les concepts d'une part, et les enjeux d'autre part, évolueront. Les choses migreront, il doit donc y avoir historisation du système (ce qui veut dire en clair, pouvoir consulter l'interface, ou les données telles qu'elle se présentaient à telle date arbitraire depuis l'instauration du schéma).

–4.1.1 Une démarche d'amenée aux principes du schéma

Un déroulement de type état des lieux, ou analyse de l'existant, puis définition des objectifs que doit atteindre le schéma, enfin modélisation du système et réalisation, pour présenter des étapes satisfaisant un esprit analytique, ne peut se correspondre à une chronologie exacte.

Un simple effort d'inventaire des acteurs (en premier lieu, les producteurs, administrateurs de données) montre qu'ils sont en très grand nombre, mal connus du point de vue de leur administration de leurs données susceptibles d'entrer dans le schéma. Il est facile d'en oublier, et d'ignorer des sources de données dont les uns et les autres disposent.

L'exercice est possible en principe. Mais les changements arrivent plus vite que l'inventaire ne converge, de façon accrue en 2008 avec une refonte en profondeur des services déconcentrés de l'État — sans par ailleurs que les directives des services centraux sur les systèmes d'information n'aient pu se coordonner à cette nouvelle donne.

Chaque acteur a potentiellement un grand nombre de gisements de données, certains ne sont plus entretenus, ou ont radicalement changé de support au cours du temps. À cause de cette péremption, leur inventaire interne est difficile, il relève de la plongée dans les archives voire de la recherche historique, sauf dans le cas rare où subsiste une personne qui a en mémoire l'ensemble de l'histoire, et qui en tient les éléments soigneusement archivés.

Il se dessine donc un ordre de priorité ou de facilité dans la publication des gisements de données de chaque acteur selon les principes du schéma directeur. Chacun est le seul à pouvoir trancher cette question, en dernier ressort, en fonction de l'enjeu que représente la publication, en fonction de l'évaluation de la difficulté à publier (ce qui peut impliquer de remanier profondément l'organisation technique — et humaine, dont en dernier ressort tout dépend), en fonction des capacités de travail qui peuvent être mobilisées.

Chacun a en propre à fournir un travail conséquent pour expliciter comment tous ses gisements de données se placent dans une perspective de migration vers des services d'information.

En premier lieu, chacun est appelé à faire un bilan propre dans lequel il délimite ce qu'il souhaite, doit ou peut publier prioritairement. C'est le processus qui est engagé, et qui va durer dans le temps.

On a donc un classement :

–4.1.1.1 D'un mode de gestion en monolithes (des menhirs)

Historiquement, l'informatisation a concerné les parties bien individualisées, dont la modélisation ne posait guère de problème, parce qu'elle était déjà faite et s'imposait d'elle-même. L'exemple-type est celui des cartes grises, dans une préfecture. C'est ce genre de noyaux autour desquels on cherche depuis le milieu des années 1990 à établir un système d'information complet. Évidemment, ces noyaux peuvent être un peu durs à croquer.

D'autant que depuis, on s'est attaqué à l'informatisation de tâches, organisées autour d'une fonction, beaucoup plus complexes. L'inventaire des données a par exemple révélé le cas du logiciel \gsp\ dans les \ddaf, qui permet d'entretenir et d'analyse financièrement l'ensemble d'un service d'eau ou d'assainissement d'une collectivité, dans le but d'aider à sa gestion, et notamment à la négociation de contrats d'affermage.

-4.1.1.2 À un mode de gestion décomposé en couches (une tour à étages)

Une décomposition en stades successifs, bien individualisés, assurant l'acquisition de données, puis des transformations successives, avec à chaque fois des modes de présentation en entrée et en sortie bien définis (selon un contrat), et se prêtant à un traitement automatique. Aux stades de la publication (la mise en accès public), on s'appuie sur des services informationnels fondés sur les technologies de l'Internet et du Web, et, le cas échéant, sur des géo-services.

À un stade antérieur, les services gestionnaires migrent vers des modes d'administration qui se conforment à cette décomposition (et aux principes énoncés plus haut).

-4.1.1.3 Avec de multiples stades intermédiaires

[Décrire un ou deux exemples, situés diversement le long de cet axe.]

-4.1.2 L'adoption d'une démarche d'inventaire permanent des composantes du schéma

-4.1.2.1 L'inventaire des gisements de données

Dans leur effort d'implication, et pour se former une idée de l'état des lieux, plusieurs acteurs primaires ont produit des croquis de leur systèmes informatiques, et des applications et bases de données qu'ils supportent. Chacun a utilisé des outils différents, essentiellement graphiques, pour schématiser à la fois leurs flux de données, leurs équipements, leurs installations logicielles. Cette présentation, pour être un bon support de prise de connaissance et de discussion, n'est pas pratique à maintenir au cours du temps, et ne permet pas (avec les outils utilisés) une liaison avec une base de données décrivant l'ensemble des composantes de chaque système d'information.

Une rapide réflexion montre que les conditions ne sont pas réunies pour proposer un dispositif unique (une application structurée autour d'un langage informatique) permettant à tous de décrire leurs systèmes de façon homogène, qui puisse être assemblée (pour faire une synthèse), facile à maintenir à jour.

Un essai infructueux a consisté de notre part à proposer une découpe simplifiée en « acteurs », « bases de données », « serveurs », « solutions techniques », présentée sous la forme d'onglets d'un classeur Excel, un des rares outils auquel tous soient assez accoutumés pour qu'il ait une chance d'être un support de langage commun. Le classeur avait été renseigné par nos soins de la totalité de ce que nous avons pu glaner au cours des réunions avec les acteurs, avec des compléments pour nombre d'informations de portée nationale.

Les deux envois que nous avons effectués auprès des acteurs n'ont recueilli aucune réponse. La refonte des systèmes déconcentrés n'aura pas facilité la tâche des services de l'État, qui étaient parmi les premiers concernés. Il est apparu aussi que l'analyse en catégories simplifiées était à la fois trop simple pour rendre compte correctement de la multiplicité des agencements de systèmes informatiques et de gestion des données, et trop complexe pour pouvoir être compréhensible au travers d'une série d'onglets dans un classeur Excel. Cette enquête n'a donc pas fonctionné.

Or une enquête est... un système d'information. En l'occurrence, nous n'avons pas su en établir un de fonctionnel.

En pratique, les liens ne sont pas encore assez tissés, les relations assez en place autour de ces questions, pour que les multiples interlocuteurs aient des compétences croisées à des niveaux permettant une mise en phase

effective. C'est toute cette démarche qui est en cours, sur tous les fronts à la fois.

Au final, on propose donc un système simple, qui pourra évoluer vers quelque chose de plus élaboré, plus intégré dans le futur. Il prend en compte pragmatiquement le matériau sur lequel on est en mesure de travailler à ce stade du schéma.

–4.1.2.2 Une suggestion d'interface pour un inventaire permanent

De simples fiches textuelles, avec une série de rubriques, largement optionnelles, pour décrire chaque gisement, mais aussi chaque point d'intérêt ou question problématique.

Avec un éditeur de texte structuré, on garde la possibilité d'insérer images (des schémas) et des url. Il faut aussi un système de versionnement, pour pouvoir suivre l'évolution des documents. La solution la plus immédiate, la plus directe à mettre en service est un wiki réservé aux acteurs.

–4.1.2.3 Un exemple de fiche décrivant un gisement de données : \gsp (dans les \ddaf)

Le sigle \gsp\ signifie gestion des services publics.

Mis en œuvre par les directions départementales de l'agriculture et de la forêt depuis les années 1990, dans le cadre de prestations d'ingénierie publiques (rémunérées) assurées auprès de collectivités qui leur en font la demande.

[Contexte sur l'importance croissante de ce type de mission : peut paraître superflu, mais permet une mise en perspective des questions en jeu sur l'environnement et l'économie de l'eau et l'assainissement, qui donnent du sens à toutes les données que recèle ce logiciel. Les missions de \gsp\ exercent un rôle de régulation sur les marchés de délégation de services publics, comme décrit ci-dessous.]

Le logiciel est assis sur une base de données servant à analyser financièrement et techniquement le fonctionnement de toutes les composantes d'un service public d'eau ou d'assainissement. Cette analyse permet de suivre l'exécution de contrats de délégation (ou d'affermage) de ces services par les collectivités qui en ont la charge, auprès d'opérateurs qu'elles chargent de leur gestion. Le travail permet aussi, initialement, de préparer et négocier ces contrats.

On peut notamment en passer par un inventaire arbitrairement détaillé de tous les équipements, réservoirs, stations d'épuration, mais aussi pompes, vannes, etc., avec leur état de vieillissement, d'amortissement, descriptions, photos, leur situation administrative (droit immobilier, foncier) et géographique.

Seule une étude approfondie du modèle des données permet de bien comprendre la signification de chacune d'entre elles. Cette étude ne consisterait typiquement pas à lire le dictionnaire de la structure des tables, avec leur description, mais à prendre en main le manuel d'utilisation, et à suivre les sessions de formation sur le suivi de la gestion des services publics dispensées dans les écoles spécialisées du ministère de l'agriculture, comme l'Engées. Par ailleurs, pour pratiquer cette activité de suivi, une solide connaissance de base dans les techniques et l'économie des services d'eau et d'assainissement est indispensable. En un sens, la compréhension du modèle des données, et donc de la signification que peut revêtir chacune des variables qu'on y trouvera, requiert ce même arrière-plan de connaissances. [On voit donc ici nettement une limite à la possibilité effective d'explicitier auprès du public la sémantique d'un jeu de données.]

Le logiciel est réputé et apprécié par ses utilisateurs pour la précision avec laquelle il permet de cerner son objet, en l'occurrence, exercer un métier de conseil et contrôle. Cependant, les services variant considérablement en importance et en complexité, et la demande des collectivités en matière de négociation et suivi des contrats de délégation s'étalant sur une large gamme, des utilisateurs sont amenés, pour leur travail, à **faire recouvrir** à d'assez nombreuses variables des **significations très différentes** de

celles d'autres utilisateurs, comme ils le constatent entre eux lors de rencontres d'échange. Cette situation est appropriée au travail réalisé à partir de ces variables, l'absence de dépendance **sur ce plan** entre les utilisateurs fait que ces discordances ne sont pas problématiques (sur un plan global, tous les contrats sont, en revanche, sur un même marché et sont donc en un sens liés économiquement).

La base de données correspondante est potentiellement extrêmement riche, avec des possibilités intégrées de représentation géographique. Cependant, elle ne concerne que les services des collectivités qui ont fait la demande de cette mission d'assistance auprès des \ddaf, et du fait même de sa richesse et de sa complexité, ce jeu de données est délicat à publier, sans parler des questions de confidentialité qui peuvent s'y attacher.

[Donner une idée des collectivités présentes, par rapport au potentiel.]

–4.2 . Réduire les écarts sémantiques, du processus classique à une interface d'interrogation

Classiquement, « l'approche sémantique » des données se faisait sans qu'on la remarque expressément, d'elle-même en quelque sorte, à travers les contacts progressifs avec l'administration et le guidage qui s'ensuivait. Au passage, la personne intéressée apprenait le vocabulaire en usage pour parler administrativement de ce que sa langue spontanée lui faisait exprimer autrement. Non sans quelques douleurs parfois, comme dans l'apprentissage de toute langue étrangère.

De nombreux processus y concourent. Par exemple, la loi fait obligation à une autorité administrative de réacheminer une demande qui lui est mal adressée, indiquant probablement qu'il est souvent très difficile pour le public d'y trouver son interlocuteur. On compte sur la culture des services et le sens du service collectif pour aiguiller les demandes, suppléer leurs mauvaises formulations, apporter des réponses pour permettre au demandeur de mieux se situer dans l'univers qu'il interroge.

Le problème central est bien un problème de communication. Il s'agit de franchir un mur de langage. Et se placer au stade de franchir un mur de langage est ce qui caractérise le fait de faire un **système d'information**, plutôt que seulement un **système de données**. Un système d'information à proprement parler intègre le stade de l'interprétation que font les individus chaque fois qu'ils entrent des données ou qu'ils en prennent connaissance.

Si l'accès à l'information ne se fait plus qu'à travers l'interface d'un système d'information automatisé, ce mécanisme d'approche doit désormais être explicite, formalisé, et réalisé dans l'interaction d'un utilisateur avec le système informatique. Toute cette mise en contexte doit être rédigée, lisible sous une forme similaire à un livre, un point unique d'entrée dans le système, avec un certain nombre de caractéristiques : sa consultation est grandement facilitée par la présence d'une table des matières (et un bon ordonnancement du plan), et celle d'un index très étendu qui permet de réduire une grande partie des différences de langage entre l'utilisateur et les contenus de données, en renvoyant un terme courant à une notion telle qu'elle est nommée techniquement. Les différences plus profondes, qui remettent en cause tout un ensemble de notions à la fois, ne peuvent être réduites qu'à la lecture d'un texte rédigé, qui a cette propriété singulière : un lecteur peut en commencer la lecture, comprendre chaque mot, et achever la lecture en ayant intégré des mots nouveaux, ou avec de nouvelles acceptions pour des mots qu'il pratiquait déjà.

–4.3 . L'interface d'exploration, le référentiel

L'**interface** doit viser à respecter deux propriétés très importantes.

Elle doit être aussi **unitaire** que possible, c'est-à-dire présenter toujours le même visage. Cela signifie idéalement qu'un seul et même écran permet de faire des requêtes, et de consulter, explorer les réponses disponibles, tout au long de l'interaction. La multiplication des écrans donne l'impression de changer d'interlocuteur, rend impossible de suivre le « fil de la

discussion », d'aller au bout de sa requête. Un écran unitaire est une très bonne présomption d'une interface très réussie.

Elle doit donner en permanence une **vision d'ensemble** des données disponibles. On peut appeler cela, dans un sens figuré, être « cartographique ».

On n'est pas dans la situation de l'espace immense de tout le web, qu'en pratique on peut considérer comme ouvert ou infini, et dont il n'est jamais question d'avoir une vue d'ensemble. On est plus dans l'idée d'un **répertoire** construit, soigné et contrôlé d'objets bien qualifiés, et on peut donc toujours connaître et représenter la totalité.

À chaque instant, l'interface doit montrer **la portion** sur laquelle on est en train de se concentrer au travers d'une requête, immergée **dans l'image de la totalité** disponible.

—4.3.1 L'atlas et le catalogue, un index et glossaire général, pour une interface intégrée

L'**atlas** est le lieu d'accès aux cartes de l'environnement de la Corse. Pour faire une carte électronique, il faut faire le geste de choisir quelles sources cartographiques y figureront. Le **catalogue** contient en principe le moyen de trouver quelles données sont disponibles, pour faire des cartes notamment, mais pas seulement, sur l'environnement en Corse. On parle de ces deux entités de façon distincte parce que l'interface de l'information sur l'environnement en Corse se présente de cette façon.

Mais un utilisateur, pour faire une carte, consulter des données, fait en principe une requête sur un espace donné (en Corse), une période donnée, et des catégories données. Il a besoin d'une interface permettant de spécifier cela, et de faire converger les catégories (métadonnées) qu'il utilise vers les termes utilisés spécialement dans le système d'information de l'environnement en Corse.

Actuellement (2008) par exemple, le catalogue fait mention des Azi ; en regardant les couches correspondantes, on sait qu'il s'agit de zones inondables, mais sans explicitation du sigle. Côté atlas, on cherche des couches pertinentes qui sont regroupées selon une tout autre nomenclature ; elle diffère aussi de celle page 8 de la note de cadrage, et de la liste page 18 (celle-ci partage cependant un certain nombre d'entrées avec les « lots » du catalogue).

Il faut absolument faire un **index général**, qui regroupe toutes ces entrées et y adjoigne les mots courants que l'on peut s'attendre à voir invoquer sur ces sujets. C'est plus un index au sens de celui que réclame un livre savant, malgré tout le soin que l'on peut apporter au chapitrage, malgré la présence d'un sommaire. On observe en effet que, quel que soit le soin apporté à la rédaction d'un ouvrage, il est impossible de confiner en un et un seul endroit le traitement de toutes les notions qui y sont abordées. C'est une autre façon d'illustrer le fait évoqué plus haut, qu'il est impossible d'établir un référentiel absolu des notions : elles prennent existence dans un certain contexte, et aucun contexte ne recouvre tous les autres, bien qu'on puisse dans une certaine mesure expliquer à l'un ce qu'est le monde d'un autre, apprendre à un pêcheur à connaître un peu le monde du baigneur, du botaniste, etc.

La richesse des données sur l'environnement obligera, sans effacer les classements « réglementaires » des entrées, à construire un index de ce type, si on veut se rendre vraiment accessible à un large public. Le travail d'indexation est un travail long, difficile, qui réclame un gros effort de concentration, et dont quelqu'un doit jouer le rôle de rédacteur en chef, pour lui assurer une cohérence indispensable à sa qualité. Il est notamment alimenté par le travail de longue haleine sur la nomenclature que les partenaires primaires du schéma ont entamé et doivent entretenir en permanence.

Voici ce que dit Isabelle Boydens, auteur de *Informatique, norme et temps* des questions de construction du référentiel, d'abolition des homonymes et

synonymes, de maîtrise de la sémantique : « La construction d'un thésaurus de descripteurs (dans les domaines juridiques ou techniques, par exemple) peut nécessiter plusieurs années-homme et l'on sait, quand on travaille dans ce domaine, que la constitution des champs sémantiques, le traitement des ambiguïtés de langage, des homonymes et des synonymes sont des tâches jamais pleinement achevées. » On a vu aussi plus haut que dès que l'on couvre une période de temps un tant soit peu étendue, on est confronté au caractère vivant des nomenclatures, comme toute langue en usage. Isabelle Boydens poursuit : « Si l'on s'adresse à des dirigeants d'entreprise ou à des gens qui sont quotidiennement confrontés à de telles questions, il me paraît plus efficace de leur dire la vérité : les difficultés auxquelles vous êtes confrontés ne sont pas triviales ; certaines d'entre elles sont irréductibles et l'harmonisation du référentiel est un objectif jamais pleinement atteint vers lequel toutefois il faut s'efforcer de tendre, en vue de contribuer à l'efficacité d'une entreprise ou d'une organisation. »

Une troisième propriété de l'interface d'exploration pourra alors jouer un rôle capital pour avancer dans ce processus.

Il s'agit de créer un retour d'information, une rétro-action qui informe si les données d'indexation sont propres ou pas à amener le public (et aussi n'importe quel partenaire primaire, même s'il est réputé mieux informé) à ce qu'il recherche. On remarquera au passage que si des milliers de personnes utilisent l'index à la recherche de données, du point de vue de l'administrateur du schéma, s'il n'a aucun retour d'information sur ces requêtes, c'est comme si les données d'index n'étaient pas utilisées. Il n'a aucune idée de leur valeur. Il est aussi l'un des plus mal placé pour les tester lui-même, puisqu'il est probablement celui qui connaît le mieux le corpus qui est indexé. Comment se perdre dans une ville que l'on connaît comme sa poche, pour l'avoir sillonnée dans tous les sens depuis son enfance, pour en avoir vu construire les rues nouvelles, une à une ?

Une méthode consiste à tracer systématiquement les errements de consultation, de requêtes des utilisateurs, et leur permettre à tout instant, dans chaque consultation de tirer une alarme, pour dire en langage naturel ce qu'ils cherchaient, et qu'ils n'ont pas trouvé. Idéalement, le formulaire par lequel ils feraient cette appel au secours présenterait une compilation de leurs requêtes, et des jeux de données successifs consultés, qu'ils pourraient annoter. L'administrateur aura aussi la trace vierge. Si le rédacteur en chef est présent à ce moment, il serait aussi en mesure d'apporter une réponse en direct, en ouvrant un dialogue avec l'utilisateur. Les vertus supérieures du dialogue pour résoudre les écarts de compréhension et converger rapidement vers une réponse satisfaisante ne sont pas à prouver. Conserver la trace de ces dialogues sera aussi une information précieuse pour comprendre comment améliorer l'index et l'interface en général.

Cette disposition permettra de faire évoluer progressivement l'interface, d'après la réalité et non d'après des idées plus ou moins préconçues ou préfabriquées.

L'index sera évidemment complété par les fiches textuelles, dont on a essayé de traduire l'esprit dans un exemple (gsp), qui jouent en quelque sorte le rôle d'un glossaire. Il importe d'éviter de multiplier les objets d'information, sous peine d'égarer à force de proposer des guides différents. Les données, avec leurs métadonnées structurées, les fiches, comme métadonnées ouvertes et capables d'explicitation, et un index doivent suffire à couvrir tous les besoins.

L'idéal est que la recherche dans l'index donne des références, fasse converger vers le vocabulaire stabilisé du schéma, mais illumine aussi les entrées concernées dans leur arborescence de présentation (si c'est une arborescence), ou dans une table à plusieurs entrées (présentées comme un hypercube), ou toute représentation qui ferait une sorte de carte des données.

La même interface doit pouvoir permettre aussi facilement d'utiliser les facilités de cartographie dynamique qu'elle intègre, et de transférer les

services sources qui y sont visualisés, si possible aussi avec leur configuration courante, vers un outil dont disposerait un utilisateur sur son poste de travail. Il y aurait ainsi une continuité parfaite entre le catalogue de services (avec gestion des droits) et le service intégré de visualisation et exploration cartographique.

—4.3.2 Ces référentiels sont des schémas de communications, des langues, elles sont vivantes, elles dérivent

Elles sont appelées à dériver au fil du temps. Les structures des relations qu'entretiennent les personnes, les contrats, les objets en jeu dans les relations évoluent avec le temps, parfois sur des périodes très courtes. Et les nomenclatures dérivent. Seules les nomenclatures qui ne servent pas, c'est-à-dire les langues mortes, ne dérivent pas.

On ne peut jamais fixer une sémantique, la maintenir compréhensible pour toujours. Elle se maintient, autant que le contexte qui l'a vu naître et qui l'entretient se maintient.

C'est pour cela que l'entretien de l'archive des données du schéma est un défi, le prochain qu'il faudra relever : Il faudra s'accommoder de cette dérive, et être capable de lancer des requêtes dans le passé.

C'est pour cela aussi que l'inventaire ne peut qu'être permanent. Le système doit être pensé comme en remaniement constant, et permettant pourtant une continuité dans le travail. Il faut que dans sa conception le système soit capable d'assimiler des évolutions.

Date: 2009-02-02 12:44:35 CET

vincent douzal

Philippe Lemoisson