

**Jean-Jacques Girardot**

Maître de conférences en Sciences Économiques  
[jjg@mti.univ-fcomte.fr](mailto:jjg@mti.univ-fcomte.fr) , + 33 6 13 33 29 58

**Adresse professionnelle**

Maison des Sciences de l'Homme Claude Nicolas LEDOUX  
Université de Franche-Comté - 32 rue MÉGEVAND - F-25030 BESANÇON cedex

**Résumé** : Cette communication actualise nos réflexions sur la fonction éditoriale des systèmes d'intelligence territoriale. L'informatique et les systèmes d'informations visent à accroître la communication entre les êtres humains. Un système d'intelligence territoriale constitue un instrument au service des acteurs qui ambitionnent de maîtriser le développement équitable et durable de leur territoire. Depuis l'irruption de la microinformatique nous avons conduit des recherches pour concevoir des logiciels conviviaux, puis pour modéliser la structure d'un système d'information territorial au service d'un réseau d'acteurs territoriaux et plus récemment pour intégrer les fonctions d'analyse des données. Ces dernières nous ont conduit à considérer aujourd'hui le système d'intelligence territoriale sous l'angle d'une chaîne d'édition numérique animée par un partenariat d'acteurs territoriaux au service d'une communauté territoriale.

**Abstract** : This communication makes an updating of our thinking about the editorial function of the systems of territorial intelligence. Computer science and information systems aim at increasing the communication between the human beings. A territorial intelligence system is an instrument at the service of the actors who ambition to control the development the equitable and sustainable development of their territory. Since the irruption of data processing, we led research activities to design friendly software, then to model the structure of a territorial information system that is at the service of a network of territorial actors, and more recently to integrate the data analysis functions. They led us to presently consider the territorial intelligence system from the prospect of a digital edition chain, which is animated by a partnership of territorial actors at the service of a territorial community.

**Mots clés** : intelligence territoriale, systèmes d'information, Systèmes d'Information Géographique (SIG), analyse des données, édition numérique, gouvernance territoriale

**Keywords**: territorial intelligence, information systems, Geographic Information Systems (GIS), data analysis, digital edition, territorial governance.

Cette communication résume les résultats de recherches financées par le Conseil Régional de Franche-Comté, la Direction Régionale de la Recherche et de la Technologie (DRRT), la Direction Régionale de l'Industrie, de la Recherche et de l'Économie (DRIRE) et par le Fond Européen de Développement Régional dans le cadre de l'Institut des Sciences et Technologies de l'Information de l'Université de Franche-Comté.

## La fonction éditoriale des systèmes d'intelligence territoriale

Un système d'information est un ensemble structuré de bases de données qui rassemble un ensemble défini d'informations et des fonctions qui permettent d'établir des relations entre ces informations, de définir le sens de ces relations ou, mieux encore, de les mesurer. Dans un Système d'Information Territorial (SIT), les informations concernent un territoire, c'est-à-dire un espace géographique et sa communauté. Elles sont référencées dans l'espace et dans le temps. Des fonctions d'analyse statistique et spatiale visent à aider les acteurs territoriaux à produire des connaissances sur ce territoire.

Ainsi un SIT permet :

- De collecter et de regrouper des informations ;
- De les analyser selon des protocoles scientifiques, notamment sous l'angle spatial ;
- D'interpréter les résultats de ces analyses ;
- De représenter les informations et les résultats ;
- D'élaborer des scénarios décisionnels ;
- De gérer et d'évaluer les politiques, les programmes, dispositifs, projets et actions qui résultent des décisions.

Un système d'intelligence territoriale est un système d'information territorial au service d'un partenariat d'acteurs territoriaux qui souhaitent développer une gouvernance démocratique au service du développement durable.

- Il favorise le partage des informations au sein d'un partenariat de développement territorial ;
- Il instrumente l'analyse coopérative des données et l'interprétation participative des résultats ;
- Il introduit la participation citoyenne dans le processus de prise de décision ;
- Il fournit aux acteurs les informations utiles pour élaborer des projets, puis pour les gérer et les évaluer.

Il exploite le potentiel que l'informatique propose pour rassembler des informations et les partager, pour favoriser le partenariat des acteurs à partir de l'analyse coopérative des informations, et pour accroître la participation des citoyens en améliorant leur information et leur accès à l'information. Il respecte les contraintes que le développement durable, particulièrement la participation et le partenariat, impose aux protocoles de traitement et de publication de l'information.

Cet article montre comment nous avons convergé, à partir de la notion de convivialité vers celle de chaîne éditoriale, via l'accessibilité des systèmes d'information et l'intégration des fonctions d'analyse de l'information. Cette évolution est liée à la diffusion des technologies de l'information, mais aussi au souci permanent de les rendre accessibles

au plus grand nombre sans préjudice de la qualité de l'information.

Le développement de l'informatique, des systèmes d'information, des réseaux numériques, puis de l'édition numérique via Internet nous ont permis de concevoir et de réaliser progressivement des solutions qui se sont avérées utiles pour les systèmes d'information territoriale. L'informatique a d'abord permis de constituer de grands ensembles de données et de les analyser avec des méthodes statistiques quantitatives. Elle a rendu possible l'application des méthodes statistiques qualitatives comme l'analyse factorielle ou la classification [BENZECRI, 1982]. Les bases de données se sont développées, puis les systèmes d'informations. Des informations, conservées sur des supports différents, ont été numérisées afin d'être stockées massivement dans un même support au sein duquel elles perdaient leur lisibilité. Internet a popularisé le partage des informations, leur traitement collaboratif et la publication massive à l'échelle planétaire.

Nous proposons de résumer l'historique de nos recherches concernant les systèmes d'intelligence territoriale, les résultats actuels et nos perspectives.

Docteur en sciences économiques, embauché dans un laboratoire de mathématique, statistique et informatique au sein d'une faculté des lettres et sciences humaines, j'ai développé dès les années 70 des solutions informatiques "conviviales" pour diffuser des protocoles analytiques issus des sciences de la matière et de la nature parmi les chercheurs en sciences humaines et sociales.

Grâce à une collaboration avec des chercheurs en informatique, l'action de recherche SITRA, Systèmes d'Information Territoriaux pour les Réseaux d'Acteurs, a permis de compléter l'analyse statistique de données par des outils d'analyse spatiale et de cartographie.

ICASIT, Intégration de Chaîne Analytique des Systèmes d'Information Territoriaux, a engagé l'actualisation et l'évolution des logiciels d'analyse statistique et spatiale des données développés précédemment.

Ces travaux ont fait l'objet d'une communication lors du colloque ICTAMI à ALBAC (Roumanie) en septembre 2006 et feront prochainement l'objet de deux présentations à Taiwan en novembre 2007.

Ce faisant, nous avons pris conscience de l'importance de la fonction éditoriale de cette chaîne analytique orientée vers l'édition en ligne

des résultats. Dans le contexte des systèmes d'intelligence territoriale, cette chaîne éditoriale concerne la production de documents numériques qui doivent être partagés au sein d'un partenariat d'acteurs. Elle constitue également un work flow qui ne peut pas être organisé uniquement en fonction des protocoles techniques d'analyse des données, mais aussi de manière à permettre la participation au sein du partenariat mais surtout, au-delà, au sein de la communauté territoriale. Pour cette raison, nous les dénomons Systèmes Communautaires d'Intelligence Territoriale (SCIT). Approfondir et instrumenter la fonction éditoriale des SCIT est notre perspective de recherche actuelle.

## 1. LE DEVELOPPEMENT DE LOGICIELS “CONVIVIAUX” D'ANALYSE DE DONNÉES.

Il est important de remonter aux logiciels qui constituent à présent les modules de base des SCIT. Ils se fondent sur une notion qui constitue un fondement de la fonction éditoriale des SCIT : la convivialité.

Le logiciel ANACONDA - ANalyse CONviviale des Données, a été publié en 1981 pour diffuser les méthodes qualitatives d'analyse des données (analyse factorielle et classification) dans les sciences humaines et sociales.

PRAGMA, logiciel de dépouillement quantitatif de questionnaires a été réalisé en 1991 pour faciliter l'élaboration des tableaux de données en vue de l'analyse statistique qualitative de données.

La notion de convivialité, fondamentale dans leur conception, impliquait déjà les idées d'économie, de simplicité dans l'utilisation, d'accessibilité, de partage et de coopération. Développés au sein du groupe de recherche “Techniques Nouvelles en Sciences de l'Homme” la conception de ces logiciels a tenu compte des spécificités propres :

- Aux sciences humaines et sociales (SHS) ;
- Aux usages pluridisciplinaires ;
- Aux collectes systématiques et collectives de grands ensembles de données ;
- À la complémentarité avec l'analyse spatiale.

### 1.1. ANACONDA

Développé en 1981 [GIRARDOT, 1981], son principal objectif était de fournir une solution économique et accessible pour l'analyse des données. Il s'agit de méthodes multicritères qui permettent d'analyser des individus statistiques décrits par un ensemble de caractères multidimensionnels. L'analyse factorielle détermine les facteurs structurels, ou tendances, d'un ensemble volumineux de données. La classification en

dissocie les principales classes. Ces méthodes impliquent d'imposants calculs qui imposent l'usage de l'ordinateur, à la différence des tableaux quantitatifs que l'on peut réaliser “manuellement” [BENZECRI, 1982]. Pendant les années 70 ces calculs étaient réalisés dans des centres de calculs à un coût important pour les laboratoires SHS. Dès que le nombre de caractères excédait cent ou que le nombre d'individus dépassait mille, il fallait “monter” à PARIS. Or les ensembles de données dépassant mille individus étaient fréquents en géographie. ANACONDA a permis d'analyser des ensembles de données jusqu'à cent caractères et mille individus sur un Apple II (48k). Ces limites ont rapidement été repoussées.

ANACONDA combine l'Analyse Factorielle des Correspondances (AFC) de Jean-Pierre BENZECRI avec la Classification Ascendante Hiérarchique (CAH) de Maurice ROUX<sup>1</sup> [1985]. Il est basé sur un algorithme de diagonalisation de matrice conçu par Jean-Philippe MASSONIE programmé initialement par Xuan LUONG.

Destiné aux chercheurs en SHS, ANACONDA a été conçu comme un logiciel gratuit et simple à utiliser pour des utilisateurs sans qualification informatique. La microinformatique a permis de simplifier l'utilisation de l'analyse des données. Il était possible de faire une analyse de données sur son ordinateur personnel. ANACONDA présentait une interface très simple comparée aux logiciels informatiques. Il était interfacé avec les logiciels bureautiques courants. Plutôt que de programmer un module d'édition des données, ANACONDA a été conçu pour importer les données sous la forme d'un tableau de données<sup>2</sup>, constitué avec un traitement de texte ou un tableur et enregistré dans un fichier en format texte<sup>3</sup>. Il est totalement automatisé pour un usage courant. Il applique un principe popularisé par Jean-Pierre FÉNELON [1981] : «avons-nous besoin d'une thèse en thermodynamique pour conduire une automobile ? ». Il ne nécessite pas d'autres connaissances en mathématiques que celles qui sont utiles à la compréhension des résultats. Les choix sont paramétrés avec des valeurs par défaut qui correspondent à 99% des utilisations. Il suffit d'indiquer le nom du fichier contenant les données à

<sup>1</sup> Basée sur la distance du Chi2 et sur l'agrégation par les moments centrés d'ordre deux.

<sup>2</sup> Le tableau de données présente les individus en ligne et leurs caractères en colonnes. La première ligne est réservée pour les codes d'identification des caractères et la première colonne pour les identificateurs des individus.

<sup>3</sup> Suffixe txt. Chaque ligne du tableau est séparée de la suite par le caractère “return” (ASCII 13). Dans chaque ligne, chaque colonne est séparée de la suivante par le caractère de tabulation (ASCII 11).

analyser pour que le calcul des contingences, l'analyse factorielle des correspondances et la classification ascendante hiérarchique soient exécutées. Le logiciel NUAGE simplifie ensuite l'analyse et l'interprétation des résultats grâce à une animation en trois dimensions.

Mais surtout, ANACONDA est adapté à la spécificité méthodologique majeure des SHS où l'analyse s'intéresse principalement aux individus par rapport aux caractères. Les méthodes expérimentales provenant des sciences de la matière concentrent l'intérêt sur les rapports entre les variables en neutralisant les individus. Plus que dans les sciences de la nature, caractérisées par la variété, les SHS se singularisent par la diversité des individus et par l'individualité. Le principal intérêt de l'analyse qualitative tient à sa capacité d'étudier cette diversité. Nous avons choisi l'analyse factorielle des correspondances qui permet d'analyser non seulement les rapports entre les variables, mais aussi les rapports entre les individus, et les rapports entre les individus et les variables. La classification permet d'approcher la diversité des individus en déterminant les principales classes d'individus et leurs profils de caractères. ANACONDA permet d'associer les deux approches par la représentation de la classification dans l'espace factoriel.

Ainsi, la convivialité d'ANACONDA repose :

- Sur l'automatisation des fonctions dont la connaissance n'est pas indispensable à la compréhension des résultats ;
- Sur un paramétrage des choix avec des valeurs par défaut couvrant les utilisations courantes;
- Sur une interface avec les standards bureautiques les plus populaires ;
- Sur une adaptation aux spécificités des sciences humaines et sociales.

ANACONDA a été lauréat du concours Apple de la Pomme d'Or, série recherche et développement, en 1981. Nous avons suivi sa diffusion pendant une dizaine d'années. Lorsque nous avons dû arrêter ce suivi, faute de temps, elle dépassait mille exemplaires. ANACONDA a été utilisé dans des domaines variés. Dans la recherche scientifique sur l'analyse des paysages en géographie et sur la structure et la dynamique des établissements archéologiques. Dans divers domaines de recherche appliquée et de recherche-action : études de la structure des collections muséales et des publics de manifestations culturelles, diagnostics de la variété et de la complexité des besoins de publics défavorisés et évaluation des actions visant à l'amélioration de la situation de ces publics, etc.

## 1.2. PRAGMA

PRAGMA se présente comme un logiciel de dépouillement d'enquêtes par questionnaire orienté vers le traitement quantitatif. Il a été initialement développé en 1991 [GIRARDOT, 1991] pour réaliser les opérations de codage qui sont nécessaires :

1. Pour effectuer des tris quantitatifs simples et croisés ;
2. Puis, pour constituer le tableau de données en vue de l'analyse qualitative des données qui sera exécutée par ANACONDA.

La conception de PRAGMA répondait en outre à de nombreux besoins qu'il a été possible de satisfaire conjointement grâce au fabuleux potentiel intuitif et interactif du logiciel HyperCard. Le microordinateur Macintosh d'Apple a popularisé dès 1984 une interface graphique qui a considérablement simplifié l'utilisation des ordinateurs. Avec HyperCard, il est devenu possible de programmer des applications intuitives proches des pratiques de travail.

### 1.2.1. Une utilisation intuitive

PRAGMA affiche chaque question sur l'écran avec une présentation similaire à celle qu'elle a sur le questionnaire papier. On saisit une réponse en cliquant sur la modalité correspondante dans la liste des modalités qui figure à l'écran. Dans le cas d'une question ouverte, on saisit sa valeur comme on le fait habituellement avec un traitement de texte. On visualise à tout moment la liste des réponses faites par un individu en cliquant sur son code. On peut alors revenir sur une question pour consulter les fréquences des modalités ou pour établir un index des valeurs permettant de consulter les individus qui ont fait les mêmes réponses. Pour coder une question, il suffit de sélectionner des valeurs proches puis de constituer une nouvelle modalité en quelques clics. On regroupe de la même façon plusieurs modalités en une seule. Les réponses sont totalisées en temps réel au niveau de chaque question : il suffit de consulter, et éventuellement d'imprimer, l'écran d'une question pour disposer du tableau de comptage correspondant, que ce soit pour l'ensemble des individus ou pour une sélection. Les tableaux simples, ou croisés, sont enregistrés dans des fichiers au format texte. Un tableur (comme Excel) permet de les lire, de compléter les calculs et de dessiner des types variés d'histogrammes.

PRAGMA s'est affirmé comme un logiciel simple à utiliser. La saisie est devenue une opération réalisable par toute personne sachant utiliser un traitement de texte. Les chercheurs, ainsi que les professionnels, ont pu réaliser les opérations de

codage, de recodage, de sélection et de tri statistique, puis compléter l'exploitation et la communication des résultats à l'aide des outils bureautiques usuels.

### 1.2.2. Un outil collectif.

PRAGMA a été conçu pour un usage collectif, pour mutualiser et partager les données en amont, et pour analyser collectivement les résultats des traitements informatiques en aval. Avec la diffusion de la microinformatique, le cercle des utilisateurs de l'informatique s'est considérablement élargi et la taille des bases de données a fortement augmenté. Au début des années 80, des ensembles importants de données étaient collectés "sur le terrain" au moyen d'enquêtes ou d'observations. Les laboratoires de Géographie, d'Archéologie et de Chronoécologie étaient notamment spécialisés dans la collecte systématique d'ensemble de données concernant plusieurs milliers d'individus ou d'observations. La collecte nécessitait la coordination de plusieurs équipes. Les nouveaux utilisateurs de l'informatique n'étaient pas des informaticiens, ni même des utilisateurs expérimentés de logiciels. Ils travaillaient dans des environnements informatiques variés. La saisie partagée a permis de constituer des bases volumineuses conjuguant les compétences de plusieurs disciplines ou de plusieurs secteurs d'activité. L'analyse et l'interprétation des résultats sont également devenues des coopérations pluridisciplinaires et multisectorielles. Les logiciels HyperCard, puis ToolBook, ont présenté l'avantage de rassembler dans un seul fichier l'ensemble des fonctions logicielles et des données. Cela a permis de regrouper sans complication des saisies réalisées par des utilisateurs dispersés, au plus proche des lieux d'enquête ou d'observation.

Ainsi, PRAGMA a permis de conduire la collecte de données concernant plus de 2155 descriptions d'établissements archéologiques, afin de les croiser avec des bases environnementales couvrant une superficie de 1500 km<sup>2</sup> dans la basse vallée du Rhône (dans un triangle entre Lyon, Nice et Montpellier en France), dans le cadre du programme « Archaeomedes » de lutte contre la désertification des sols en Europe (1992-2000). Il a ensuite été utilisé pour établir la carte sociale de HUELVA (Espagne) avec des données multisectorielles concernant 3852 ménages ; le diagnostic du Plan Départemental d'insertion du Doubs (1994) auprès de 8897 ménages ; ou encore l'évaluation de 16 « Boutiques Solidarité » de la Fondation Abbé Pierre auprès de 5895 sans abri en France (1995). Ces bases ont été constituées par plusieurs dizaines d'équipes de chercheurs ou d'acteurs. Elles ont défini et validé un questionnaire commun intégrant plusieurs thèmes disciplinaires

ou plusieurs approches sectorielles. Ce questionnaire a été enregistré dans PRAGMA. Un exemplaire, vierge de données, a été distribué à chaque équipe. Chacune a collecté et saisi ses données dans un fichier partiel. Ces fichiers ont été regroupés en une seule base PRAGMA, identique aux bases partielles mais contenant tous les individus statistiques. Elle a fait l'objet d'une analyse globale, d'analyses partielles et éventuellement d'une confrontation avec d'autres données. Les équipes qui avaient partagé leurs données ont été associées à l'analyse et à l'interprétation des données.

La convivialité de PRAGMA repose ainsi sur la reproduction de l'environnement usuel de travail par l'outil informatique. Elle se fonde ensuite sur l'intégration de PRAGMA dans l'environnement bureautique. Elle prépare ainsi une transition vers des usages nouveaux qu'un environnement de travail habituel ne peut pas offrir.

PRAGMA met également en valeur la dimension collective de la convivialité, fondée sur les notions de partage et de coopération. Nous prenons mieux conscience à présent de ses implications.

1. Elle suppose la constitution d'une communauté d'utilisateurs unie par un projet commun.

2. Elle nécessite une organisation pour définir les tâches concourant à la réalisation de l'objectif, pour inventorier des moyens utiles, pour établir un calendrier de travail et pour mettre en œuvre une communication. Toutefois, la communauté n'est pas toujours structurée par une hiérarchie décisionnelle. La communication n'est pas canalisée de haut en bas. Elle fonctionne également de bas en haut et de façon transversale. Les droits des membres sont égaux, ce qui implique un processus de proposition et de validation des décisions.

3. Chaque membre de la communauté doit trouver son avantage dans la coopération sans préjudice pour la valeur ajoutée globale. Il est ainsi impératif de définir préalablement les droits de chacun sur l'information partagée. L'apport d'informations et le temps qui lui est consacré supposent une restitution qui intéresse les participants à titre collectif et particulier. La participation à l'analyse et à l'interprétation des résultats sont le complément du partage d'information.

4. L'amélioration de l'accessibilité de l'information, données et résultats, est une autre condition de l'implication des membres de la communauté.

5. Ainsi que le transfert de compétences scientifiques et techniques concernant les méthodes, les protocoles et les outils d'analyse utilisés.

PRAGMA devait également offrir la possibilité d'exploiter statistiquement des bases de données. La diffusion de la microinformatique dans les années 80 a également contribué au développement important des bases de données. De nombreuses entreprises et administrations ont constitué des bases de données qui pouvaient se prêter à des analyses statistiques. Toutefois elles n'avaient pas été constituées dans ce dessein. Il fallait les restructurer pour respecter les contraintes d'un traitement statistique.

### 1.2.3 Robustesse et contrôles automatiques de la qualité des données

Afin de favoriser l'accès d'un nombre croissant d'utilisateurs aux techniques d'analyse quantitative et qualitative de l'information, mais aussi pour faire face à la demande d'exploitation statistiques des bases de données, PRAGMA intègre des procédures de contrôle de la qualité des données.

Ces contrôles concernent essentiellement la forme des questions. Ils se fondent d'une part sur la spécificité du questionnaire et, d'autre part, sur la résolution des contradictions qui caractérisent ses deux fonctions complémentaires : communication et structuration des traitements.

Le questionnaire est un document interactif de collecte d'information qui structure cette dernière de façon à simplifier son traitement. Il permet de poser à tous les individus enquêtés le même nombre de questions, dans le même ordre et selon la même formulation. Lorsque les individus statistiques ne sont pas des personnes, mais des objets, les contraintes les plus importantes concernent le nombre et le sens des questions. Le sens de chaque question est établi avant de procéder à la collecte des données et il ne doit plus être modifié ensuite. Il s'agit de définir un langage commun à toutes les personnes qui seront associées à la collecte, à l'analyse, à l'interprétation et à la publication des données. PRAGMA considère que la conception et la rédaction du questionnaire constituent une opération préalable. Il permet de numériser le questionnaire une fois qu'il est rédigé. Il contrôle la forme des questions mais non leur formulation.

Communication et structuration de l'information peuvent s'avérer contradictoires. En particulier, la question ouverte, qui appelle une réponse individualisée, est préférable du point de vue de la communication. La question fermée, qui propose de choisir une réponse parmi une liste de modalités prédéfinies, est appropriée au traitement car elle permet d'opérer directement des tris statistiques, de réaliser des graphiques et des cartes géographiques, puis de préparer les analyses qualitatives. Toutefois,

PRAGMA n'est pas un logiciel normatif. Il ne contraint pas à utiliser seulement des questions fermées. Il ne fige pas la forme des questions car il considère qu'elle est évolutive : il suffit de fermer la question ouverte en codant ses réponses. Ainsi, PRAGMA autorise la saisie et l'importation de questions sous forme ouverte et sous forme fermée. Il contrôle la forme des questions afin de proposer ensuite des fonctions de codage assisté.

Les contrôles automatiques concernent :

- L'absence de réponses.
- L'unicité ou la multiplicité des réponses.

PRAGMA repère et compte systématiquement les absences de réponse en les dissociant automatiquement des situations où la question devient sans objet.

Il vérifie étroitement la saisie des questions "à réponse unique" ou "à réponse multiple", et notamment si les modalités sont exclusives ou peuvent être utilisées une ou plusieurs fois.

Les questions auxquelles il manque des réponses et celles qui acceptent plusieurs réponses sont en effet plus complexes à trier et à analyser. Dans les deux cas, le nombre total de réponses diffère du nombre d'individus interrogés. Il lui est inférieur lorsqu'il manque des réponses. Il est supérieur s'il est possible d'indiquer plusieurs réponses. Il s'ensuit des confusions selon que les pourcentages et les histogrammes sont calculés en fonction du nombre d'individus ou du total de réponses exprimées. Pour limiter ces confusions, PRAGMA calcule ces deux pourcentages, sur le bilan quantitatif global exporté sous le format "texte".

Outre le contrôle des erreurs de manipulation informatique, ces contrôles formels font de PRAGMA un outil robuste. Il a pu être utilisé dans des conditions extrêmement diversifiées, au plus près des situations d'observation, par des personnes sans qualification informatique ou statistique - ce qui ne signifie pas sans formation ni suivi. Ces expériences ont permis en retour d'améliorer la robustesse de PRAGMA, qui constitue un facteur important de convivialité.

PRAGMA complète les contrôles automatiques, par des fonctions de codage et de recodage, qui proposent une aide à l'utilisateur selon un protocole de qualification des données :

- Gestion des réponses absentes.
- Codage des questions ouvertes.
- Recodage des questions fermées.
- Sélection de questions et de modalités.
- Codage en vue de l'analyse des données.

Ces fonctions, qui demandent plus d'expertises, sont étroitement assistées pour éviter les erreurs autant que faire se peut.

La gestion des réponses absentes concerne à la fois les individus et les questions. En principe, il faudrait éliminer tous les individus pour lesquels il manque une réponse. Toutefois, le coût d'acquisition des données ou le souci de ne pas introduire des biais dans un échantillon peuvent s'opposer à une perte d'information drastique. Dans ce cas, il faut veiller à éliminer les questions et les individus pour lesquels le taux de réponses absentes est trop important.

Dans la mesure où le contrôle des réponses absentes abouti à l'élimination d'individus et de questions, il constitue la première étape du protocole de qualification des données, au terme de laquelle, ainsi qu'à chaque étape, le bilan quantitatif global pourra être aisément actualisé.

Le codage des questions ouvertes vise à les transformer en questions fermées, c'est-à-dire les réponses exprimées sous forme de valeurs en codes. PRAGMA propose simplement de regrouper des valeurs entre elles selon une nouvelle modalité. Il est également possible de coder directement une valeur dont le sens correspond déjà à une modalité existante. Ainsi, au moment de la saisie, il est possible de coder directement une réponse ou d'indiquer sa valeur en cas de doute. Elle pourra être ensuite codée soit en l'affectant à une modalité existante, soit en créant une nouvelle modalité.

Le recodage des questions fermées est une opération de regroupement similaire qui porte non plus sur des valeurs, mais sur des réponses qui ont déjà la forme de codes. Il concerne les modalités dont la fréquence est trop faible pour qu'elles soient considérées comme représentatives. PRAGMA permet de supprimer ces modalités, ou lorsque la signification des modalités le permet, de les regrouper avec d'autres modalités ou entre elles afin d'éviter la perte d'information.

Le recodage est également un moyen de sélection des questions ou des modalités en fonction du stade d'analyse, exploratoire ou approfondi, ou encore pour préparer la communication des résultats.

Ces fonctions de codage, de recodage et de sélection ont notamment été développées pour sélectionner les caractères qui constitueront le tableau booléen de données en vue de l'analyse qualitative des données. PRAGMA permet de sélectionner globalement les questions à analyser, puis de choisir et de définir les caractères à analyser à partir des modalités, éventuellement en regroupant plusieurs modalités. Ce « tableau de

données » constitue un lien important pour l'intégration de PRAGMA et d'ANACONDA. Inversement, la typologie déterminée par ANACONDA peut être importée dans PRAGMA comme une question dont les classes constituent les modalités, afin de calculer par un tri croisé les fréquences des modalités originelles dans chaque classe.

Ces opérations sont également utiles pour la communication des résultats ; pour associer les partenaires qui ont participé à la mutualisation des données à l'interprétation des résultats et pour diffuser les résultats sur Internet. Dans les deux cas, les résultats bruts de l'analyse quantitative ne sont pas publiables en l'état et doivent faire l'objet d'une sélection et de recodages spécifiques.

Nous avons insisté sur la nécessité de restituer les résultats au réseau de partenaires qui acceptent de partager leurs données. Avec PRAGMA, cette restitution s'effectue au niveau individuel dès la saisie puisque chaque partenaire peut effectuer les analyses quantitatives et qualitatives qui l'intéressent à partir de ses données avant même de les mettre en commun. Mais la restitution a surtout une dimension collective après le regroupement des données. Les partenaires peuvent participer à l'analyse, orienter la progression des traitements et enrichir l'interprétation des résultats par l'apport de données d'expérience, d'informations ou de connaissances spécifiques. Cette restitution suppose une animation fondée sur la communication des résultats.

La publication des résultats sur Internet suppose pour sa part une élaboration progressive des documents sous forme numérique. Ces documents possèdent de nombreuses caractéristiques particulières. Ils sont nouveaux dans des systèmes d'information généralement conçus pour des documents plus traditionnels. Ils sont produits de façon automatique, mais ils sont également refaçonnés par des interventions humaines en fonction d'impératifs techniques ou communicationnels. Ils sont également très diversifiés.

### 1.3. NUAGE

NUAGE est un logiciel de représentation en trois dimensions des résultats de l'Analyse Factorielle des Correspondances (AFC) et de la Classification Hiérarchique Ascendante (CAH). Il a été développé en 2003 afin de remplacer le logiciel américain MACSPIN dont le développement avait été arrêté depuis 1991. Le logiciel d'analyse graphique des données MACSPIN permettait de réaliser une représentation tridimensionnelle des résultats d'ANACONDA. Cette représentation était

nettement plus simple à utiliser que les représentations par plans sur papier. Elle était surtout beaucoup plus intuitive et permettait l'interprétation collective des résultats des analyses qualitatives en petits groupes. Mais MACSPIN pouvait être seulement utilisé sur ordinateur Apple, et il est devenu pratiquement impossible de l'utiliser après 2002 avec la généralisation du système d'exploitation « X » d'APPLE.

NUAGE, développé en JAVA par Hervé FEHNER, est un logiciel multiplateforme. Il restitue toutes les fonctions graphiques de MACSPIN, mais son interface est spécifiquement adaptée à la double représentation de l'analyse factorielle (AFC) et de la classification (CAH) et à l'interprétation de l'analyse qualitative des données. De ce fait, elle intègre automatiquement une série d'opérations qui devaient être réalisées manuellement auparavant. Elle mémorise également les résultats des analyses et des représentations. Malgré un travail important sur l'interface, elle produit essentiellement des documents de travail. La réflexion doit s'orienter à présent sur la production de documents éditables.

## 2. SITRA.

L'action de recherche "Systèmes d'Information Territoriaux pour Réseau d'Acteurs", a été conduite entre 2002 et 2004, afin de modéliser des systèmes d'information territoriaux (SIT) conviviaux, accessibles en ligne à des réseaux d'acteurs locaux du développement territorial durable en vue de développer leur capacité de mobilisation, de décision, d'action et d'évaluation. Il s'agissait notamment de définir une architecture de SIT adaptée au travail en réseau et de compléter l'analyse statistique des données avec des solutions cartographiques conviviales. Les résultats de SITRA ont été appliqués pour élaborer le système de gestion des ressources de l'ADRI (Agence pour le Développement des Relations Interculturelles, devenu à présent la Cité Nationale de l'Histoire de l'Immigration), le système de gestion des ressources de la Délégation Interministérielle à la Ville (DIV), et pour faire évoluer les observatoires territoriaux « CATALYSE ».

Le terme de réseau renvoyait à la réalité d'acteurs coopérant sur un projet commun, mais exerçant leur travail dans des organismes distincts au plan institutionnel, sans lien hiérarchique et géographiquement éloignés. Il préfigurait la notion de partenariat qui constitue un principe économique fondamental du développement durable. Ce dernier préconise le recours aux ressources locales dans un esprit de coopération plutôt que la création de nouvelles structures. Il suppose une approche pluridisciplinaire et multisectorielle. Le partenariat est un terme à présent courant au niveau de l'action

territoriale. Le programme européen d'action socio-économique « EQUAL » finance par exemple des actions conduites par des consortiums appelés "partenariats de développement". Suite au développement de la coopération au niveau de l'analyse de données grâce à la convivialité de PRAGMA, les enjeux de SITRA concernaient toujours le partage de l'information et son exploitation coopérative. Il s'agit toutefois maintenant de mobiliser de façon cohérente des informations extrêmement variées référencées dans l'espace et dans le temps, pour qu'elles fassent l'objet d'une analyse statistique ou d'un autre usage. L'action SITRA réunissait des chercheurs en sciences humaines et sociales et en informatique, alors que nous avons jusqu'ici réalisés les logiciels d'analyse de données en interne. Nous voulions utiliser le potentiel d'Internet et cela impliquait des connaissances et des compétences plus étendues en réseaux informatiques, en modélisation des systèmes d'information et en bases de données.

SITRA a permis :

- d'engager l'intégration des fonctions d'analyse statistique et spatiale en concevant des outils cartographiques conviviaux ;
- de définir une procédure de gestion globale de l'ensemble des documents, en dépit de leur variété ;
- de définir un système de droits approprié à une gestion partenariale ;

Le terme de "systèmes d'information territoriaux" (SIT) renvoie à un projet à la fois plus ambitieux et plus simple que celui de "système d'information géographique" (SIG). Les SIG permettent de référencer géographiquement les informations, d'effectuer des analyses spatiales, de les analyser, et de cartographier les informations et les résultats. De ce point de vue, un SIG constitue le module de base d'un SIT. Toutefois, l'expérience du projet « ARCHAEOEMEDES » et des observatoires « CATALYSE » avait montré que les bases de données des SIG ne pouvaient pas gérer de façon cohérente la variété des documents impliqués, ni la diversité des acteurs, ni le flux documentaire. En outre, leur usage restait coûteux et complexe pour des performances médiocres en édition numérique sur Internet. Si ce coût et cette complexité pouvaient être justifiés pour des analyses spatiales spécialisées, ce n'était pas le cas pour réaliser des analyses spatiales simples et pour éditer des cartes, fonctions pour lesquelles existait une offre gratuite plus conviviale. Il suffisait que l'architecture du SIT s'inspire de la structure d'un SIG pour intégrer ces fonctions. ARCHAEOEMEDES utilise un SIG pour effectuer des analyses spatiales, alors que le mal dénommé SIGVILLE dessine des cartes à la demande concernant des indicateurs socio-économiques publics appropriés à l'observation des territoires et à l'évaluation des dispositifs sur



l'ensemble des territoires de la politique de la ville (environ 700 zones infra et inter communales) à l'aide d'un outil cartographique en ligne (ALOVMAP). La manipulation est extrêmement simple puisqu'il suffit de sélectionner la zone et la donnée à représenter. Cette solution conviviale est également utilisée pour la cartographie des diagnostics et des indicateurs territoriaux dans les observatoires CATALYSE en utilisant notamment des fichiers en format texte pour cartographier les résultats issus de PRAGMA et d'ANACONDA. SITRA a ainsi poursuivi l'intégration des logiciels statistiques avec les nouvelles fonctions cartographiques. Il a engagé la réalisation d'une version libre et multiplateforme d'ANACONDA en Java, et un prototype libre en ligne de PRAGMA, epragma, en PHP/MySQL.

Un SIT gère l'accès en ligne d'une large variété de documents en réseau sécurisé Extranet ou sur Internet. SIGVILLE gère également la documentation relative aux actions de la politique de la ville (contrats, plans, rapports). C'est un élément du site de ressources en ligne de la politique de la ville : ouvrages, périodiques, revues, circulaires, rapports, cartes, conventions, tableaux de chiffres, graphiques, fiches d'expériences, liens de sites partenaires ... Cette variété est encore enrichie par l'approche pluridisciplinaire et multisectorielle qui caractérise l'action territoriale. Le site de l'ADRI propose un système simplifié de métadonnées afin d'améliorer l'accessibilité des documents via Internet. Alors qu'il proposait encore plusieurs recherches en fonction du type de document, SITRA a modélisé pour les ressources de la politique de la ville un SIT qui référence et qui gère de la même façon l'ensemble des documents référencés et numériques. On accède ainsi avec un seul mot-clé à l'ensemble des documents relatifs à un thème, quels que soient leur support d'origine, leur format numérique, leur type ou leur fonction. Un SIT est caractérisé par la production de nouveaux documents à partir d'autres documents. La gestion des documents y concerne également le référencement des nouveaux documents et la gestion du flux analytique.

SITRA a également engagé la modélisation d'un système de sécurité adapté aux réseaux partenariaux, en proposant une procédure qui s'est généralisée depuis. L'administrateur propose un login et un mot de passe modifiables par l'utilisateur. SITRA a repris de PRAGMA l'organisation des individus par groupes. Cette notion offrait la possibilité de rassembler les individus renseignés par une même structure ou habitant une même zone territoriale. SITRA a introduit la différenciation des accès en fonction des groupes, ce qui permettait, en définissant les groupes en fonction d'un organisme, par exemple,

de limiter aux utilisateurs de cet organisme le droit d'accéder aux individus appartenant à ce groupe. Les travaux sur la sécurisation des SIT ont abouti à approfondir le lien entre métadonnée, document et information, en fonction de l'actualisation de l'information. Nous pouvons illustrer cette réflexion comme suit. Nous pouvons constituer pour la représentation cartographique d'indicateurs un tableau dont les colonnes sont les indicateurs et les lignes les unités géographiques à cartographier. Nous pouvons considérer que ce tableau est un document et lui attacher des métadonnées, comme la date de mise à jour, et des droits. Ce sera insuffisant si chaque indicateur ou chaque zone peut être actualisé indépendamment par des utilisateurs différents. Dans ce cas, nous devons attacher les métadonnées et les droits à la colonne ou la ligne correspondante du tableau. Si nous sommes susceptibles de modifier indépendamment chaque information, nous devons attacher des métadonnées à chaque information du tableau.

SITRA a ainsi permis d'établir une première série de spécifications pour la réalisation des SIT et de les expérimenter avec des réalisations importantes comme SIGVILLE.

### 3. ICASIT.

L'action ICASIT "Intégration de la Chaîne Analytique des Systèmes d'Information Territoriaux", a poursuivi SITRA de 2004 à 2006 en mettant l'accent sur la notion de chaîne analytique qui débouchera finalement sur celle de chaîne éditoriale.

SITRA a posé les bases d'une chaîne analytique, intégrant les fonctions logicielles au sein du SIT, de la collecte des données avec PRAGMA jusqu'à la publication en ligne des résultats, principalement des cartes. ICASIT a poursuivi l'effort d'intégration des logiciels d'analyse avec deux objectifs pratiques :

1. Réaliser des versions intégrées, libres et multiplateformes, des fonctions logicielles du SIT.
2. Mettre en ligne ces fonctions.

NUAGE a été développé en version multiplateforme Java pour représenter en 3D les résultats d'ANACONDA. Les spécifications d'une version multiplateforme de PRAGMA ont été établies. Les fonctions cartographiques ont été étendues avec la représentation des flux.

Une version bêta de PRAGMA en ligne, epragma, a été expérimentée dans le cadre des observatoires des migrations de l'ACCEM.

Cet effort d'évolution et d'intégration s'est réalisé et se poursuivra avec le souci constant d'améliorer

l'accessibilité des outils logiciels de plusieurs façons :

- En utilisant et en réalisant des logiciels libres, largement ouverts à la communauté des développeurs.

- En réalisant des versions multiplateformes, fonctionnant sur tous les ordinateurs, sur tous les systèmes d'exploitation et indépendantes de leurs évolutions.

- En automatisant, comme cela avait été engagé avec ANACONDA, toutes les fonctions dont ni l'expertise, ni la maîtrise technique, ne sont indispensables à la compréhension des résultats. Il reste en creux les étapes opérationnelles et décisionnelles qui requièrent une intervention des utilisateurs via l'interface homme-machine. La poursuite du traitement requiert de nouvelles informations, généralement transmises à l'aide de documents, ou de décisions transmises sous forme de paramètres conditionnant le traitement.

- En intégrant de l'expertise. Des valeurs par défaut sont habituellement attribuées aux paramètres par des experts. L'utilisateur dispose d'une proposition qui correspond soit aux usages courants, soit à des conseils experts, et qu'il peut modifier. Par exemple, dans SIGVILLE, l'utilisateur peut demander la représentation d'une carte en indiquant la zone et l'indicateur. Le choix des classes et des couleurs de représentation de chaque indicateur sont paramétrés par des experts. Ils intègrent une expertise adaptée à chaque indicateur d'une part, et à la représentation cartographique, d'autre part.

- En mémorisant les décisions des utilisateurs. Les démarches d'évaluation et d'observation impliquent souvent la répétition de décisions d'une période sur l'autre. Les paramètres sont utilisés pour mémoriser les décisions.

La définition d'un guide multisectoriel de diagnostic et d'évaluation, coordonné au sein de la CAENTI, vise à proposer une sélection de questions et pour chaque question un protocole de traitement en fonction de l'expertise et de l'expérience des partenaires de la CAENTI.

Les recherches conduites au sein de SITRA sur l'intégration des fonctions d'analyse de l'information ont mis l'accent sur le flux de documents rassemblés et produits au sein du SIT, de la collecte à la publication des données. Ainsi que nous l'avons souligné précédemment un SIT référencie et numérise une grande variété de documents dont certains restent peu usuels (tableaux et graphiques statistiques, cartes ...), élargie par la production de documents numériques. ICASIT a posé les bases de la modélisation du flux documentaire en débutant

l'inventaire des documents utilisés (input) et produits (output) par la chaîne analytique. L'objectif est de définir un système de métadonnées qui décrivent la source des informations et les traitements réalisés, si possible de façon automatique. Dans la continuité de l'effort de simplification initié par SITRA, ces recherches visent à intégrer le standard « DUBLIN CORE », notamment les normes actuellement en développement pour l'information géographique. C'est également une condition pour que les métadonnées constituent des vecteurs de communication avec d'autres systèmes.

ICASIT a mis en relief la fonction éditoriale des SIT et leur dimension communautaire.

L'intégration du processus analytique sur Internet est conditionnée par la publication en ligne des résultats. La production des connaissances et la documentation des informations s'inscrivent dans un "workflow" d'édition numérique scientifique et technique qui organise la continuité des fonctions techniques de la collecte des informations à la publication des résultats, via le partage des données, l'analyse et l'interprétation des résultats, la rédaction et la documentation des documents et leur validation.

Grâce à l'édition numérique sur Internet les SIT proposent un formidable potentiel pour le développement durable. Les outils de coopération en ligne sont indispensables au partenariat. L'édition numérique permet d'informer la communauté territoriale, d'assurer la visibilité de son identité et de développer la participation au moyen de services interactifs. Ce potentiel fait des SIT des systèmes d'intelligence territoriale qui proposent des solutions concrètes d'instrumentation du partenariat et de la participation. Toutefois ce potentiel est initialement maîtrisé par les experts. L'accessibilité de l'information territoriale et la convivialité de ses outils de traitement sont les conditions pour libérer ce potentiel au service de la l'intelligence collective de la communauté territoriale. Leur ambition est de constituer des systèmes communautaires d'intelligence territoriale (SCIT).

Cette orientation suppose la catégorisation des utilisateurs des SCIT – a priori : experts, acteurs, partenaires et communauté, et l'étude de leurs usages. Quelles fonctions nécessitent une expertise ? Quelles sont les fonctions transférables aux utilisateurs ? Selon quelle pédagogie et avec quel accompagnement ? Comment organiser le partage des informations territoriales, puis la coopération au niveau de leur analyse et de leur édition ? Quelles sont les fonctions des acteurs, du partenariat et de la communauté territoriale ?

Comment associer la communauté au moyen de services interactifs participatifs ? Cette modélisation suppose une adaptation des étapes de l'analyse des données et de l'interprétation des résultats, ainsi que du flux éditorial, aux objectifs de communication, d'animation du partenariat, de participation et de gouvernance de la communauté territoriale.

## CONCLUSION

Le développement de la convivialité des méthodes et outils génériques de large applicabilité d'analyse de l'information en direction des chercheurs en sciences humaines et sociales débouche à présent sur l'ouverture des SIT au partenariat des acteurs territoriaux et à la participation de la communauté territoriale.

Cette évolution reflète celle de la convivialité qui faisait à l'origine essentiellement référence à la simplification de l'accès individuel aux technologies informatiques. Elle a pris ensuite une dimension collective dans le cadre du partage des informations et de son traitement coopératif. Elle englobe alors progressivement l'accessibilité à l'information et à ses outils d'analyse, d'interprétation et d'édition.

Les recherches sur les systèmes communautaires d'intelligence territoriale (SCIT) concernent ainsi à présent :

- La modélisation de leur architecture et la définition de leurs spécifications fonctionnelles en fonction des logiques propres à l'analyse, à la documentation, à l'édition et à la communication dans le cadre du partenariat et de la participation ;
- La modélisation des contenus, des documents et du workflow éditorial en intégrant l'ensemble des interventions humaines individuelles et collectives ;
- Les spécifications des métadonnées ;
- La description des protocoles de traitements ;
- L'évaluation des usages ;
- La modélisation des processus décisionnels, de la gouvernance, du partenariat et de la participation au sein de la communauté territoriale.

Le terme de communauté territoriale ne fait pas ici référence à une dimension biologique, historique ou religieuse. Elle désigne un ensemble d'individus unis par un projet commun. C'est également le sens employé en informatique pour désigner une communauté d'utilisateurs. Ce recouvrement de sens permettra d'étudier les conditions dans lesquelles la communauté territoriale peut constituer la communauté des utilisateurs du SCIT. Il faut tenir compte du fait que si l'analyse d'informations peut constituer un projet pour les utilisateurs du SCIT : c'est une étape intermédiaire au service d'un objectif de développement durable

pour la communauté territoriale, ou plus simplement pour les acteurs locaux qui souhaitent projeter, conduire et évaluer des actions de développement local.

## BIBLIOGRAPHIE

BENZECRI, J.-P., 1982: "Histoire et préhistoire de l'analyse des données", Dunod, PARIS.

BENZECRI, J.-P. et F., 1976: "Pratique de l'analyse des données", 2 tomes, Dunod, PARIS.

FENELON, J.-P., 1981: "Qu'est-ce que l'analyse des données ?", Lefonen, PARIS.

GHIGLIONE, R., MATALON, B., 1978: "Les enquêtes sociologiques. Théorie et pratique", Armand Colin, PARIS.

GIRARDOT, J.-J., 2005: "Concepts, principes et outils de la méthode CATALYSE", in : *3e colloque international du Réseau Européen d'Intelligence Territoriale*, LIÈGE (Belgique), 20 et 21 octobre 2005, 5 p. <URL:<http://mti.univ-fcomte.fr/reit>>

GIRARDOT, J.-J., 2005: "Integration of data analysis tools within territorial information systems.", International Conference on Theory and Applications of Mathematics and Informatics, ALBAC (Romania), September, 15-17.

SANDU, D., 2005: "Dezvoltare comunitară. Cercetare, practică, ideologie.", Iami, Editura Polirom.

VAN DER LEEUW, S., FAVORY, F., GIRARDOT, J.-J., 2004: "The archaeological study of environmental degradation. An example from Southern France.", in CH. L. Redman & al., "*The Archaeology of global change*", Smithsonian Institution, WASHINGTON, 112-129.

GIRARDOT, J.-J., 2000: "Principes, Méthodes et Outils d'Intelligence Territoriale. Évaluation participative et Observation coopérative", in : *Conhecer melhor para agir melhor, Actes du séminaire européen de la Direction Générale de l'Action Sociale du Portugal*, EVORA (Portugal), 3-5 mai 2000, DGAS, LISBONNE, décembre 2000, 7-17.

GIRARDOT, J.-J., 1987: "Expertise du logiciel MACSPIN d'analyse graphique de données.", Sciences et Vie Micro, n°37, 100-103, PARIS (France).

GIRARDOT, J.-J., 1982: "ANACONDA, système conversationnel d'analyse des données.", *Cahiers*

*du SURF*, 1, nouvelle série, 37-74, Université de Franche-Comté, BESANÇON (France).

MASSELOT C., 2006: "Systèmes d'information territoriaux et politiques sociales : quand l'observation territoriale s'empare du net", in : 15ème Colloque de la Société Française des Sciences de l'Information et de la Communication, BORDEAUX (France), mai 2006.

MASSONIE, J.-P., 1990: "Pratique de l'analyse des correspondances.", collection "*Techniques Nouvelles en Sciences de l'Homme*", Annales Littéraires, BESANÇON (France).

ROUX, M., 1985: "Algorithmes de classification", Masson, PARIS.

Sélection de liens :

Portail Internet du centre de ressources de la délégation interministérielle à la ville (DIV, PARIS, France) <http://i.ville.gouv.fr>

Système d'information géographique de la politique de la ville (PARIS, France)  
<http://i.ville.gouv.fr/sigville>

Observatoire des migrations ODINA en Asturies (Espagne) <http://www.odina.info>

Observatoire des migrations OPASI de SIGÜENZA (Espagne) <http://mti.univ-fcomte.fr/siguenza>

Observatoire communautaire OPTIMA de SERAING (Belgique) <http://www.optima-obs.org>

Observatoire Intercommunal INTEGRA (Belgique)  
<http://www.integraplus.be>

Agence pour le Développement des Relations Interculturelles ADRI (PARIS, France)  
<http://www.adri.fr>,  
<http://www.alterites.com>,  
<http://www.histoire-immigration.fr>