

Est-il utile de réaliser des logiciels dédiés ? Réflexions à partir du logiciel EMA et de l'exemple de réalisation d'un Système National d'Information sur les Prix au Bénin

Jean-Paul Minvielle

Introduction

Mon propos dans cette note sera d'analyser rapidement la place tenue dans un système national d'information sur les prix par un instrument informatique tel EMA. Contrairement à des SGBD (Systèmes de Gestion de Bases de Données) polyvalents classiques, l'adoption d'un progiciel du type d'EMA n'est pas neutre. En effet, fléché vers des utilisations bien spécifiques, un tel instrument intègre de fait des modèles méthodologiques de création et de gestion des données, et d'analyse de l'information. Cette mise à disposition d'un système élaboré (et développé en coopération avec ses utilisateurs) peut être perçue comme positive (en permettant, sans délais et incertitudes de développement, le démarrage immédiat des enquêtes et des analyses) ou au contraire contraignante (en interdisant les configurations trop "personnalisées").

1 Le contexte de l'expérience au Bénin

L'expérience de conception et de mise en oeuvre d'un système national d'information sur les prix au Bénin a débuté en janvier 1991, à la suite de demandes formulées auprès du PNUD, durant l'année 1990, par différentes institutions (Ministères du Plan, des Finances, du Commerce, BCEAO, FMI etc.). La synthèse du problème posé était ainsi faite par le PNUD (fiche de projet MCAT/90.04) :

"Le problème de la disponibilité de statistiques fiables relatives aux prix se pose avec une grande acuité en République du Bénin. Pour pallier cette insuffisance, certaines structures utilisatrices de données statistiques se sont organisées à leur niveau pour saisir les données nécessaires à leurs besoins. Cette situation a créé dans le pays un cadre inorganisé de collecte d'informations statistiques dans lequel se côtoient les structures utilisatrices, productrices ou productrices utilisatrices, employant des méthodes de collecte différentes et souvent peu sûres. Ainsi, pour le même produit, deux structures recueillent des statistiques de prix différentes.

En vue d'harmoniser les procédés adoptés par les différentes structures intervenant dans la collecte et le traitement des données statistiques relatives aux prix,

il est nécessaire et urgent d'organiser un séminaire regroupant toutes les structures concernées, séminaire au cours duquel seront débattues les questions de méthodologies en vue de leur harmonisation dans le souci d'une production régulière des indices de prix en République du Bénin."

Le séminaire national sur "l'harmonisation des méthodologies en matières d'enquêtes de prix" que j'étais chargé d'animer, tenu à Cotonou du 22 au 25 janvier 1991, a marqué le démarrage effectif de l'expérience.

2 Quel est l'apport d'un système tel EMA ?

2.1 La disponibilité immédiate d'un instrument approprié et testé

La gestion informatisée des données fait trop souvent l'objet de confusions entre des pratiques très diverses et très inégalement performantes. En particulier, en ce qui concerne l'assimilation abusive du stockage des données à la constitution de véritables bases ou banques de données. Il n'est donc pas inutile de préciser les acceptions à accorder à chacun de ces termes : stock, base et banque (de données).

Le stockage informatique des données est une pratique très courante consistant à introduire dans des logiciels divers (tableurs, SGBD, logiciels de cartographie etc.) un certain nombre de données de base. En général, ce stockage est organisé de façon à répondre aux besoins de traitements et de résultats identifiés au départ par le gestionnaire direct (en particulier publication de documents). Les logiciels utilisés (ou les combinaisons de logiciels) peuvent parfois permettre des traitements complexes (graphiques, statistiques etc.), mais nécessitent en général de solides bases en informatique. La structure des fichiers est conçue en fonction de ces objectifs précis et, souvent, ne se prête pas à d'autres formes d'interrogation ou de combinaisons de données. Ces méthodes de stockage informatisé sont généralement très personnalisées et rendent difficile leur utilisation par des tiers et leur diffusion à l'extérieur. Les modalités d'accès et de traitement sont souvent complexes et hors de portée des demandeurs d'information. Bref, il s'agit le plus souvent d'outils de spécialistes, spécifiquement configurés par leur créateur et gestionnaire, et hermétiques aux utilisateurs extérieurs (auxquels ils ne sont d'ailleurs jamais "confiés").

Une base de données informatisée se définit non seulement par ses capacités de stockage mais surtout par ses fonctionnalités d'accès et de traitement des données. Ces fonctions devraient être accessibles à tous, sans connaissances techniques préalables, dans de bonnes conditions de simplicité et de sécurité d'utilisation. En effet, si l'on souhaite une utilisation réelle des données rassemblées, il est obligatoire que celles-ci soient manipulables facilement et efficacement. Nous sommes là très loin des pratiques habituelles en la matière, le recours aux stocks informatisés de données étant le plus souvent conditionné par le déplacement de l'utilisateur sur le lieu même du stockage, ainsi que la disponibilité et le bon vouloir du technicien ayant en charge leur gestion. Le résultat est connu, ces données sont le plus souvent inutilisées en dehors des besoins directs du service concerné.

Une banque de données comporte toutes les fonctionnalités précédentes, complétées par sa capacité à être alimentée de l'extérieur par des bases de données différentes. Il s'agit, en quelque sorte, d'une fédération de bases de données (par exemple, bases du Ministère de l'Agriculture, du Commerce, du Plan etc, rassem-

blées au sein d'une banque de données unique).

L'aboutissement ultime du système est de permettre aux utilisateurs de gérer sans intermédiaires leur demande d'information à partir d'un accès direct, sur leur lieu de travail, à la banque de données. Jusqu'à présent, pour des raisons évidentes de normalisation et de fonctionnalités, la plupart des stocks de données informatisées existants ne permettent pas ces gestions décentralisées.

L'utilisation d'un logiciel normalisé, directement dédié à ces tâches, permet en premier lieu d'économiser des temps de développement dont le coût n'est pas toujours clairement perçu. Les exemples abondent de développements de programmes spécifiques, très personnalisés, menés pendant plusieurs années par des experts sur leurs lieux d'affectation. Apparemment gratuits, ces programmes s'avèrent en fait extrêmement onéreux : au temps passé (rémunération de l'expert), ce sont des investissements de plusieurs dizaines, voire centaines de milliers de dollars qui sont ainsi consentis. Les résultats ne sont d'ailleurs pas toujours conformes aux espérances. Pour en demeurer au cas du Bénin, on peut ainsi citer le cas de l'enquête budget-consommation de 1986/87, toujours non dépouillée en 1991 après l'échec du développement d'un programme de traitement informatique spécifique. Au Bénin toujours, la Base de Données Agricoles (BDA) du Ministère du Développement Rural demeure inachevée, après qu'un expert y ait consacré plusieurs années de programmation avant de changer d'affectation en 1988. Il est peu probable que ces investissements importants puissent un jour être valorisés. Partout en Afrique de l'ouest, et certainement ailleurs, de tels exemples abondent. Qui plus est, la plupart des programmations réussies sont elles-mêmes vouées à des utilisations limitées, tant dans l'espace que dans le temps. La raison essentielle en est l'inexistence de guides d'utilisateur et de manuels techniques du programmeur, les efforts du développeur s'étant souvent limités à la seule écriture du code informatique, l'absence de maintenance et de développements. Ainsi beaucoup d'applications fonctionnelles, hermétiques aux utilisateurs extérieurs, ne peuvent être transférées sur d'autres sites ou continuer à être utilisées après que leur auteur ait quitté le service concerné. Enfin, l'utilisation d'un système normalisé permet d'envisager la réalisation de systèmes nationaux ou régionaux cohérents, basés sur son implantation dans différents sites. Cette normalisation permet le transfert de données informatisées homogènes entre services et leur utilisation décentralisée, chaque institution ayant ainsi un accès direct à la banque de données nationale ou régionale. Des systèmes d'échanges ou d'abonnements pourraient être proposés.

2.2 L'accès à un capital d'expériences :

Contrairement à des logiciels non dédiés, un décisionnel du type d'EMA intègre, de fait, un certain nombre de modèles méthodologiques résultant de la mise en commun, de la capitalisation des techniques et des expériences de traitement et d'analyse de l'information conçues et mises en oeuvre par ses différents utilisateurs durant plusieurs années. Ces modèles touchent tous les domaines, depuis les modalités de codification des produits et des zones géographiques jusqu'à la publication des tableaux finaux. Dans certains cas, des améliorations très simples de présentation ou de combinaison des données peuvent induire des gains importants dans la signification de l'information finale. La conduite du développement avec les utilisateurs permet un affinement permanent des fonctionnalités et une adaptation sans cesse meilleure aux besoins.

Par certains aspects cependant, cette facilité peut se transformer en con-

traintes, parfois mal acceptées. Cette rigidité constitue cependant une garantie méthodologique non négligeable. J'en donnerai ici un exemple, tiré de l'expérience en cours au Bénin.

Dans les relevés de prix effectués jusqu'ici par l'INSAE¹, les définitions exactes des produits étaient données par les enquêteurs sur le terrain et indiquées sur des fiches de collecte manuscrites. Il en résultait une apparente souplesse à ce niveau des enquêtes qui, dans un premier temps, fit considérer EMA comme moins adapté que des gestionnaires de bases de données polyvalents permettant, par exemple, de saisir en janvier le prix du pantalon en jean coréen de marque X, en février celui du jean américain de marque Wrangler, etc. Cette solution de facilité au niveau de la saisie et de la constitution de la base de données entraînait une accumulation d'informations hétérogènes, impossibles à traiter ensuite avec la rigueur nécessaire. Dans l'exemple cité, un parmi tant d'autres, la comparaison du prix du jean coréen de sous-marque en janvier avec celui du jean américain de grande marque en février équivalait, toutes proportions gardées, à comparer le prix d'une Renault 18 avec celui d'une Mercedes 190. On imagine aisément la fiabilité d'un indice de prix calculé sur de telles bases. Avec l'utilisation d'EMA, basé sur des codifications préalables et des définitions très précises des produits, ces approximations n'étaient plus possibles, garantissant par là même l'adéquation des données saisies aux traitements qui en seront faits. Dans ce cas, le choix d'EMA a rendu incontournable une réflexion méthodologique approfondie sur les définitions des produits qui, si elle n'avait pas été menée à ce stade, aurait conduit plus tard à des impasses analytiques. On connaît d'ailleurs les solutions généralement apportées à ce genre de difficulté : des agrégations sont effectuées, permettant de produire des mercuriales et des indices dont la fiabilité est souvent (à juste titre) mise en question. C'est vraisemblablement là une des raisons pour lesquelles les relevés effectués par l'INSAE ne sont plus publiés depuis 1987, bien que la collecte sur le terrain se soit plus ou moins poursuivie.

2.3 Des fonctionnalités facilitant la transformation des données en informations

Je l'avais déjà exposé ailleurs², information et données de base ne doivent pas être assimilées. L'information procède de l'agencement, de la combinaison des données. Une série de prix, par exemple, a dans l'absolu une valeur informative extrêmement limitée, sinon nulle face à certaines interrogations. Elle ne deviendra signifiante que dans la mesure où elle pourra être croisée avec des données complémentaires permettant de l'explicitier : niveaux des stocks, des productions, caractéristiques de l'offre et de la demande, flux etc.

Ces possibilités de combinaisons de données, et donc de création d'information, sont conditionnées par trois préalables :

- que les variables et les paramètres retenus aient été choisis de manière à se compléter dans un schéma d'analyse cohérent.

¹Institut de la Statistique et de l'Analyse Economique du Ministère du Plan du Bénin

²CT. J.P Minvielle :

"Crise et information in Politique Africaine n° 39 : Ajustement structurel et politique alimentaires en Afrique au sud du Sahara".

- "Le prix, variable fondamentale des espaces céréaliers régionaux en Afrique de l'ouest", Lomé, 6 au 11 novembre 1989

- que ces données soient suffisamment homogènes pour pouvoir être *effectivement combinées et que ces combinaisons puissent être effectuées de manière simple, sûre et rapide*. En d'autres termes, qu'elles soient gérées dans des banques de données opérationnelles.
- que les méthodes de formalisation et de présentation de l'information soient adaptées, claires et parlantes.

Si le premier préalable apparaît évident, le second n'en est pas moins important. De la même manière que, sauf contrainte majeure, on n'utilisera pas une automobile nécessitant chaque jour 1/4 d'heure de mécanique avant de fonctionner, on n'aura pas recours à des stocks de données au mode d'emploi complexe et à la manipulation fastidieuse. Le troisième enfin, qui découle en partie des deux premiers, n'est pas le moins important : une carte ou un graphique sont souvent plus explicites que des tableaux de chiffres.

Après avoir rappelé que :

“L’information régionale a pour vocation essentielle de permettre les comparaisons, de déterminer les situations relatives des zones les unes par rapport aux autres (déficits, excédents, avantages comparatifs etc.). Ne pourront, bien évidemment, être comparées que les informations comparables, ce qui sous-entend des homogénéités :

- conceptuelles sur les types de prix (gros, détail), de produits, de marchés etc ;
- méthodologiques sur les fréquences des relevés et de la production des résultats, la fiabilité des données etc ;
- techniques sur les logiciels employés, les structures de fichiers, les méthodes de transfert des données etc.”

J’étais amené à conclure, à l’issue du test, que :

“Malgré la masse de données disponibles, l’utilisateur souhaitant effectuer des comparaisons régionales entre ces trois pays (Mali, Niger, Togo), se verra réduit à ne pouvoir traiter que le seul prix du maïs à la consommation au Mali et au Niger (ceci était dû à l’hétérogénéité des types de produits et de prix). Si, par une organisation remarquable, il lui était possible de se faire communiquer les disquettes des fichiers informatiques concernant ce produit, il disposerait de deux fichiers à structure différente, gérés par des logiciels différents (SPSS au Mali, DbaseIII au Niger ?). Sous réserve qu’il connaisse les deux logiciels, il passerait probablement beaucoup plus de temps à tenter de normaliser ces deux fichiers pour les fondre en un fichier unique utilisable par l’un ou l’autre des deux logiciels (ou un troisième...), qu’à ressaisir, à partir des documents publiés, l’ensemble des données sur son système personnel. Tout laisse à penser que le travail qu’il aura alors effectué ne répondra qu’à sa propre demande, et n’aura pas d’usage collectif.”

On pourrait d’ailleurs noter que ce problème de l’homogénéité des données dans l’espace se pose aussi dans le temps. En effet, il n’est pas rare que, dans une même institution, les typologies et les structures de fichiers évoluent d’une année sur l’autre, rendant délicates, sinon impossibles, les comparaisons inter-annuelles.

Conclusion

L’utilité de la réalisation de logiciels dédiés du type d’EMA apparaît assez évidente. Cette évidence est d’ailleurs mieux perçue dans le secteur des entreprises privées que dans celui du développement : pratiquement aucun médecin ne développe son propre système de facturation, aucune PME son système de comptabilité ou de gestion des stocks etc. Cependant l’investissement (considérable) demandé par la réalisation d’un logiciel ne me semble justifié que si, outre un ciblage correct des besoins, il répond à certaines exigences de base telles la maintenance, l’assistance aux utilisateurs, les garanties de commercialisation et de développement etc.

Le logiciel EMA

La réalisation du logiciel EMA a débuté en 1987. Il est testé dans divers pays depuis 1988, dans des versions de mieux en mieux adaptées ayant permis de capitaliser les acquis et les expériences des différentes institutions ayant participé à sa mise au point.

EMA est édité, sous licence ORSTOM, par la société IXEL³. Cette association avec des professionnels permet de garantir aux utilisateurs :

- l'obtention d'un produit fiable, entièrement débuggé, conforme aux normes commerciales habituelles,
- le maintien du logiciel aux standards techniques les plus récents,
- un service après-vente et de maintenance,
- des développements du logiciel permettant de conserver la valeur de l'investissement consenti dans la création de bases de données EMA.⁴

EMA est un logiciel directement exécutable (.EXE) fonctionnant sur tous les micro-ordinateurs compatibles PC (MSDOS) disposant d'un disque dur et d'une mémoire de 640 K0. La cartographie ainsi que la représentation graphique nécessitent des cartes graphiques EGA ou VGA.

Conception du logiciel EMA

EMA a été conçu pour être l'instrument de la réalisation d'un système régional d'information pour la sécurité alimentaire. Cependant, par ses fonctionnalités, et en particulier par ses modules analytiques de cartographie et d'analyses mathématiques, il est tout à fait adapté à d'autres usages : projets de développement, usages didactiques et d'enseignement et, plus généralement, pour toutes les utilisations impliquant la gestion de données spatialisées.

EMA est organisé autour de quatre modules :

- Un module de base de données numérique sur les prix,

³ - Le module de représentation et d'analyse spatiales.

- Un module de gestion des atlas (fonds de carte et dossiers issus de la base de données cartographique),
- Un module d'analyse mathématique de séries chronologiques et de projections à court terme.

3 Un module central : la base de données sur les prix

Ce module de base de données sur les prix comporte toutes les fonctionnalités requises, accessibles par menus déroulants explicites, et donc utilisables par n'importe quel demandeur d'information, même non-informaticien :

- Paramétrage du logiciel,
- Gestion des codes,
- Gestion des fichiers. Saisies de données,
- Analyses de données (tris, recherches, statistiques),
- Editions de sorties imprimées.

4 Un premier module analytique : la cartographie

EMA étant destiné à gérer des bases de données nationales et régionales, l'approche spatiale est une de ses composantes essentielles. Qu'il s'agisse de la variable prix ou de toute autre variable intégrée dans la base de données, celles-ci peuvent être analysées à partir des cartes fournies avec le logiciel. Ces fonds de cartes, lorsqu'ils n'existent pas déjà, peuvent être réalisés à la demande.

Comme dans le logiciel CHOROSCOPE dont il est issu, ce module permet d'utiliser les cartes soit à titre purement illustratif, soit pour répondre à des questions spécifiques par le biais de croisements de variables (opérateurs logiques "ET" et "OU"):

- Représentation de la répartition spatiale des prix, de la pluviométrie, des niveaux de production etc.
- Recherche de caractères d'isomorphisme (par exemple, zones d'isoprix) ou de co-occurrences spatiales (par exemple, recherche des zones potentiellement déficitaires par la combinaison entre fortes densités de population et faibles productions).
- Recherche des zones d'opposition de tel ou tel phénomène au caractère etc.

5.1 Les fonds de carte

Ils sont numérisés à la demande et peuvent être aussi nombreux que nécessaires. Chaque fond de carte est basé sur le découpage de l'espace demandé par l'utilisateur (découpage administratif, écologique carroyage etc.). Ils peuvent comporter au maximum 9 niveaux d'emboîtements géographiques. Par exemple, le fond servant aux tests du Système Régional d'Information sur les Prix comporte 4 niveaux d'emboîtement : préfectures, régions, pays, ensemble régional regroupant plusieurs pays (en l'occurrence : Mali, Niger, Burkina-Faso et Togo). Les analyses peuvent, par focalisations successives, porter sur chacun de ces niveaux. Suivant le type de découpage retenu, le nombre d'unités primaires par fond de carte peut varier de 200 à 400.

5.2 Les dossiers cartographiques

Chaque dossier est un ensemble pouvant regrouper jusqu'à 19 séries numériques spatialisées.

Par exemple :

- Le dossier "prix du maïs en 1989" pourrait regrouper les 12 séries mensuelles des prix du maïs en 1989 dans l'ensemble régional étudié,
- Le dossier "bilans céréaliers en 1989" pourrait regrouper les niveaux de production par zone, les évaluations de population, les évolutions mensuelles des stocks et des soldes d'échanges (zones importatrices ou exportatrices) etc.

Toutes les données rassemblées dans ces dossiers sont immédiatement accessibles et utilisables, sans connaissances préalables, et les résultats des choix ou des traitements effectués sont directement représentés sous la forme de cartes, à l'échelle géographique retenue.

Ces dossiers peuvent être constitués directement dans EMA, de trois manières différentes :

- A partir des séries numériques existant dans la base de données sur les prix EMA.
- A partir de données externes transférées dans EMA.
- A partir de saisies directement effectuées dans la base de données cartographique.

Chacun de ces dossiers thématiques peut ensuite être conservé, modifié, trans-

6 Un second module analytique : analyses mathématiques et projections à court terme⁶.

Les séries temporelles de prix, comme d'ailleurs toutes séries de données chronologiques, peuvent faire l'objet d'analyses mathématiques et de projections à court terme.

Les analyses mathématiques des séries chronologiques sont basées sur la différenciation entre leurs trois composants de base : tendance, variations saisonnières et aléas. Ces analyses permettent l'identification du modèle mathématique le mieux adapté à des prévisions à court terme. Ces prévisions, présentées sous la forme de graphiques, sont calculées depuis l'origine de la série de référence, ce qui permet de visualiser immédiatement la validité du modèle retenu.

Par exemple, dans le cas d'une projection d'évolution du prix d'un produit donné en 1990, pour lequel nous disposons d'une série complète en 1988/89, la représentation graphique montrera deux courbes : la courbe réelle constatée de janvier 1988 à décembre 1989 et la courbe prévisionnelle, calculée par le modèle mathématique de janvier 1988 à décembre 1990. Cette présentation permet de vérifier, sur les années 88/89, la capacité qu'aurait eu le modèle à prévoir correctement l'évolution réelle, et donc sa validité. Suivant la qualité de la corrélation entre données réelles et données calculées en 88/89, la précision de la projection sur 1990 pourra être évaluée.

Ces séries chronologiques peuvent être constitués de trois manières différentes :

- A partir des séries numériques existant dans la base de données sur les prix EMA.
- A partir de données externes transférées dans EMA.
- A partir de saisies directement effectuées dans le module de Projections à Court Terme (PACT).

⁶Ce module d'analyse mathématique et de prévision à court terme a été réalisé par Gérard Chauvat, professeur de mathématiques à l'Université de Tours, à partir du logiciel PACT dont il est l'auteur.