

Distr.  
GÉNÉRALE

CES/SEM.47/3  
21 janvier 2002

FRANCAIS  
ANGLAIS et FRANCAIS seulement

COMMISSION STATISTIQUE et  
COMMISSION ÉCONOMIQUE POUR L'EUROPE

COMMISSION DES COMMUNAUTÉS  
EUROPÉENNES

CONFÉRENCE DES STATISTICIENS  
EUROPÉENS

EUROSTAT

Séminaire commun CEE-Eurostat sur les  
Systèmes intégrés d'information statistique  
et les questions connexes (ISIS 2002)  
(Genève, Suisse, 17-19 avril 2002)

Point I: Utilisation de la technologie du Web pour l'intégration des statistiques

## UTILISATION D'UN PORTAIL WEB POUR ACCÉDER AUX BASES DE DONNÉES DE RÉFÉRENCE COMMUNES

### Communication sollicitée

Rapport envoyé par Statistique Canada <sup>1</sup>

**Résumé :** Cet article présente un cadre architectural permettant l'intégration des bases de données de référence, dans un portail Web commun. Cette architecture porte sur l'aspect *cohérence*, plutôt que sur l'aspect fonctionnel. On présuppose que les fonctions de base des composantes existantes doivent être préservées, tandis qu'il faut offrir les nouveaux processus et services à l'ensemble de l'organisation. Le cadre proposé permet d'éliminer les incohérences dans les définitions et assure une interface de service intégrée pour plusieurs groupes d'utilisateurs. L'architecture présentée offre une méthode permettant de gérer les principales sources de référence dans une organisation statistique.

### I. INTRODUCTION

1. Pour élaborer cette architecture, plusieurs choix ont été faits selon un point de vue technologique. Ces choix reflètent une tendance généralisée au sein de l'industrie et tiennent compte du souhait des utilisateurs, qui demandent une approche basée sur le Web. L'hypothèse de base est que la plupart des services seront offerts sur les ordinateurs de bureau via un intranet d'entreprise, l'accès se faisant par un navigateur commun. Le cadre utilise aussi la technologie commerciale EAI (*Enterprise Application Integration* – intégration d'applications d'entreprise), qui est décrite plus loin dans cet article.

2. Les exemples ci-dessous qui illustrent l'architecture et ses concepts sont, pour la plupart, tirés du domaine administratif, plutôt que du traitement statistique proprement dit. Nous avons utilisé ces exemples

---

<sup>1</sup> Préparé par Mel J. Turner (Mel.Turner@statcan.ca).

en raison de leur simplicité et de leur facilité de compréhension. Nous examinons à la dernière section l'application de cette architecture aux bases de données statistiques.

3. On doit souligner ici que cet article représente une approche architecturale en vue d'une implémentation future, plutôt que la description d'un système déjà implémenté. Comme de nombreuses autres organisations, Statistique Canada est en transition vers les technologies basées sur le Web. Notre organisation doit néanmoins préserver ses investissements dans les applications et les données existantes. Nous sommes également confrontés à la disponibilité sur le marché d'un très grand nombre de nouveaux produits, chacun constituant un élément partiel de la solution, et à l'émergence, à un rythme sans précédent, de « normes » technologiques. L'architecture proposée cherche donc à instiller une certaine cohérence dans ce contexte en évolution rapide.

#### **A. Qu'est-ce qu'un portail Web?**

4. Aux fins de cet article, un portail Web est un site Web qui constitue le point d'entrée vers un ensemble d'informations et de services connexes. L'expression *point d'entrée* a une connotation de sécurité ou de contrôle d'accès, ce qui signifie que les utilisateurs doivent s'identifier pour ouvrir une session dans le portail. La nature et la disponibilité des services fournis dépendent des privilèges associés à cette identité.

5. Nous utilisons le terme « Web » dans le contexte des technologies habituellement associées à Internet ou au World Wide Web. Toutefois, cet article porte exclusivement sur les services internes qui pourraient être implémentés dans un intranet d'entreprise, au lieu d'être offerts aux utilisateurs extérieurs.

#### **B. Utilisateurs et services**

6. Pour offrir des services à des communautés d'utilisateurs disséminées au sein d'une organisation, les technologies Web constituent la solution préférée. L'avantage global pour les utilisateurs est de pouvoir obtenir des services à partir de n'importe quel ordinateur de bureau, grâce à un navigateur commun. Pour les responsables de la mise en œuvre du système, ces technologies simplifient grandement la distribution des outils logiciels, par rapport à la génération précédente de systèmes client-serveur.

7. Afin d'élaborer l'architecture présentée dans cet article, nous avons formulé certaines hypothèses de base au sujet des utilisateurs et de la gamme de services qui sont fournis par l'intermédiaire d'un portail Web, notamment :

- ?? Les utilisateurs peuvent se trouver n'importe où dans l'organisation, pourvu qu'ils aient accès à un poste de travail connecté à l'intranet de l'organisation.
- ?? Les utilisateurs sont munis d'un nom et d'un code leur donnant accès au système informatique. Les utilisateurs doivent ouvrir une session sur le portail Web pour pouvoir accéder aux services. Le navigateur Web utilise des fichiers témoins (« cookies ») de session pour contrôler les accès, ces fichiers étant configurés pour expirer après 10 minutes d'inactivité. Après ce délai, l'utilisateur doit rouvrir la session.
- ?? Les utilisateurs ont des rôles différents : employé, gestionnaire, administrateur, développeur. L'utilisateur peut personnaliser une session Web selon les privilèges accordés à son rôle, et il peut accéder uniquement aux services rattachés à ce rôle. Chaque utilisateur a un rôle par défaut qui est repris d'une session à l'autre. Au besoin, les utilisateurs doivent explicitement modifier leur rôle.

#### **C. Objectifs de l'architecture**

8. L'architecture décrite dans cet article est une *architecture de référence* de haut niveau. L'objectif d'une telle architecture est d'offrir un fondement conceptuel d'élaboration, plutôt que des recommandations spécifiques portant sur des produits ou des méthodes d'implémentation particulières. Cette architecture cherche à intégrer les applications existantes, dont bon nombre emploient des technologies différentes et

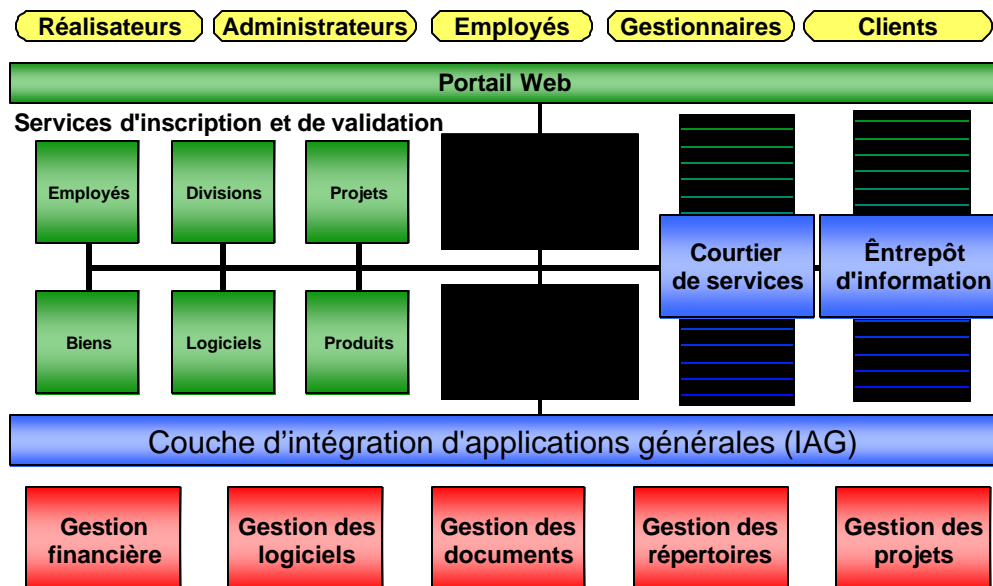
gèrent les données d'une manière non uniforme. L'architecture permet d'offrir une interface de services commune.

9. Du point de vue de l'utilisateur, il importe d'éliminer les incohérences et d'offrir une vue unifiée de l'information et des services disponibles. Dans la mesure du possible, l'intégration devrait permettre aux utilisateurs de travailler plus efficacement grâce à l'automatisation des tâches, à l'élimination de la saisie en double des données et à l'harmonisation des flux de travail entre les applications.

10. Du point de vue du développement, l'objectif est d'explorer les possibilités offertes par les nouvelles technologies commerciales, notamment par la classe de produits dite EAI (*intégration d'applications d'entreprise*) et par les infrastructures intergicielles dont elles dépendent. L'architecture touche également la question des *entrepôts de données* car, à maints égards, l'intégration de sources multiples de données, en vue de la production de rapports, présente les mêmes problèmes associés à l'harmonisation des concepts et de l'information.

11. On constate sans surprise que les métadonnées constituent la clé pour intégrer les services d'application et l'information. Dans le contexte d'un portail Web, les métadonnées de types différents et nombreux font partie intégrante du fonctionnement et de l'exploitation du système, et offrent directement aux utilisateurs des spécifications techniques et de l'information. Dans cet article, nous examinons les nouvelles exigences que cette architecture impose aux métadonnées.

## II. UN CADRE ARCHITECTURAL



### A. Aperçu de l'architecture

12. Le cadre illustré ci-dessus présente les principales composantes de l'architecture permettant d'intégrer les services administratifs types qu'on retrouve dans toute entreprise. Au bas du diagramme, les blocs représentent les applications et les bases de données existantes, souvent gérées indépendamment par différentes divisions au sein de l'organisation : ce sont ces applications et bases de données qui fournissent les services et permettent de gérer l'information administrative.

13. Le cadre traite de l'utilisation des *métadonnées* et des processus d'*enregistrement* qui offrent une vue intégrée unique aux utilisateurs, même si les bases de données sous-jacentes ne sont pas uniformes. La

combinaison des registres et des autres métadonnées ouvre la voie à l'intégration conceptuelle ou l'harmonisation des données détenues séparément par les composantes des applications existantes.

14. L'architecture sert avant tout à offrir des services à des groupes d'utilisateurs qui accèdent au portail Web. À cette fin, on utilise la composante *Courtier de services* qui a accès à un type spécial de métadonnées appelées *Règles d'entreprise*, dans le diagramme.

15. La composante d'intégration finale, dans cette architecture, est l'*Entrepôt d'information*. Cette composante remplace la myriade de modules de rapport qui caractérisent les systèmes actuels. Ceux-ci doivent être remplacés, car ils exposent l'utilisateur au manque d'uniformité et aux incohérences des systèmes sous-jacents. Un entrepôt est une base de données d'une conception spéciale, contenant des microdonnées qui sont fournies régulièrement par divers magasins de données distincts et séparés. Toutefois, les données que contient l'entrepôt doivent être complètement harmonisées avec les métadonnées visibles par l'utilisateur et avec le contenu des registres. En d'autres mots, pour assurer l'uniformité des données, il faut souvent reformater et mapper les données sous-jacentes dans un format commun.

## B. Gestion des identités

16. Pour offrir une vue unifiée aux utilisateurs, le premier point dont il faut s'occuper équivaut à la modélisation des données dans la conception d'une application : quelles sont les entités primaires faisant partie des services qu'on veut offrir? Dans l'exemple administratif que nous utilisons, il pourrait s'agir d'employés, de projets, de services, de divisions et d'autres biens de l'entreprise, qu'on doit désigner par un nom (ou une identité quelconque) quand les utilisateurs demandent de l'information ou des services. Ces identités deviennent des *noms* dans toute phrase décrivant une instance de service, par exemple : « Assigner un *nom d'employé* à ce *projet* ».

17. L'architecture du portail comporte une base de données contenant des *registres d'identités*<sup>2</sup> qui couvre l'ensemble de l'entreprise et qui ne fait partie d'aucune des applications existantes qu'on veut intégrer. Ces registres servent à vérifier l'authenticité de tout identificateur utilisé dans les interfaces de dialogue utilisateur et, par conséquent, dans toute transaction réalisée pour une application. En outre, les registres contiennent les attributs entièrement dépendants de ces identificateurs, qui sont requis pour établir des communications utiles avec les utilisateurs. Par exemple, un *code d'identification d'employé* est habituellement un numéro qui ne veut pas dire grand-chose, et qui est remplacé par le nom de l'utilisateur dans les interfaces de dialogue utilisateur.

18. Le portail devient alors le gestionnaire de ces identités au niveau de l'entreprise, bien qu'une ou plusieurs applications exécuteront également cette même fonction. Le portail offre également des *services d'enregistrement* qui mettent à jour les registres et lancent des transactions afin de synchroniser les bases de données héritées.

19. La gestion des identités illustre bien un important principe architectural au sujet des bases de données héritées sous-jacentes :

*Le portail peut avoir un accès direct en lecture dans les bases de données des systèmes hérités, mais il ne les met pas à jour directement. Plutôt, il transmet des transactions et des messages asynchrones aux applications héritées, en vue d'une action ultérieure. Les seules bases de données qui sont directement mises à jour sont celles qui appartiennent au portail, et qui sont composées de registres, de métadonnées et de règles d'entreprise.*

---

<sup>2</sup> L'identité est la base de l'intégration structurelle (qu'il s'agisse de regrouper les entités juridiques individuelles et des entités opérationnelles dans le cadre d'une entreprise, ou de regrouper des concepts de métadonnée individuels dans un système de classification).

20. Il peut y avoir de rares exceptions à ce principe de lecture seule mais, en règle générale, ce principe est requis pour assurer l'évolutivité du portail afin qu'il puisse accepter de nombreux utilisateurs et qu'il ne soit pas envahissant pour les applications héritées. En outre, cela signifie que certains services offerts par l'intermédiaire du portail ne sont pas en direct, car certaines opérations sont différées jusqu'à la mise à jour prévue des systèmes hérités.

21. Les registres du portail jouent également un rôle dans la *gestion des privilèges*, lesquels déterminent les services auxquels les utilisateurs peuvent avoir accès et l'information qu'ils peuvent voir. Nous recommandons de tenir compte du Modèle d'information des registres, proposé par Gallagher et Carnahan<sup>3</sup>. Ce modèle propose un registre complet permettant de faire le suivi des rôles et des privilèges, et incorpore les processus d'enregistrement envisagés par cette architecture.

### C. Gestion des métadonnées

22. La base de métadonnées intégrée illustrée dans le diagramme est un autre type de registre de portail. Comme les autres registres, celui-ci sert essentiellement à valider le vocabulaire des interfaces de dialogue utilisateur, quand les utilisateurs emploient le portail. Dans ce cas-ci, l'accent est mis sur les éléments, les concepts et les classifications de données, ainsi que sur leurs représentations dans le portail et les systèmes d'application qui y participent. La base de métadonnées sert également à décrire et contrôler plusieurs aspects du portail lui-même, notamment les formulaires de saisie, les labels, les listes de sélection et l'aide descriptive, tous ces éléments étant requis dans l'interface utilisateur.

23. Si la base de métadonnées peut être mise à jour dynamiquement par un utilisateur enregistré comme *développeur*, on peut utiliser directement le portail pour définir de nouveaux services d'entreprise qui peuvent subséquentement être offerts aux autres utilisateurs.

24. Dans un contexte orienté objet, la base de métadonnées définit les *types* d'objets et leurs caractéristiques connexes, tandis que les registres d'identités définissent les *instances* de ces objets et en font le suivi.

### D. Courtage de services

25. La combinaison des registres d'identité et de métadonnées qui appartiennent au portail est suffisante pour valider un large éventail de transactions destinées aux systèmes hérités d'arrière-plan. Bon nombre des conditions qui entraîneraient le rejet d'une transaction par les applications d'arrière-plan peuvent faire l'objet d'une vérification préalable et être corrigées par l'utilisateur. En outre, les métadonnées qui décrivent les formats des transactions peuvent servir à formater ou transformer adéquatement les données afin qu'elles conviennent à chaque application.

26. Le portail peut également contenir un registre de *règles d'entreprise*, lesquelles constituent un type spécial de données d'entreprise. Ces règles servent à contrôler l'ordonnement dans un portail et la façon dont les services fournis par les applications d'arrière-plan sont offertes aux utilisateurs. On trouvera des exemples plus détaillés de courtage à la section ci-dessous, *Intégration des applications*.

### E. Gestion de l'entrepôt

27. Enfin, la dernière classe de service qui pourrait être offerte par le portail consiste à soutenir l'accès et la présentation des données enregistrées dans les magasins de données opérationnelles, à l'intérieur même des applications d'arrière-plan. Dans la plupart des implémentations, l'entrepôt de données est indépendant

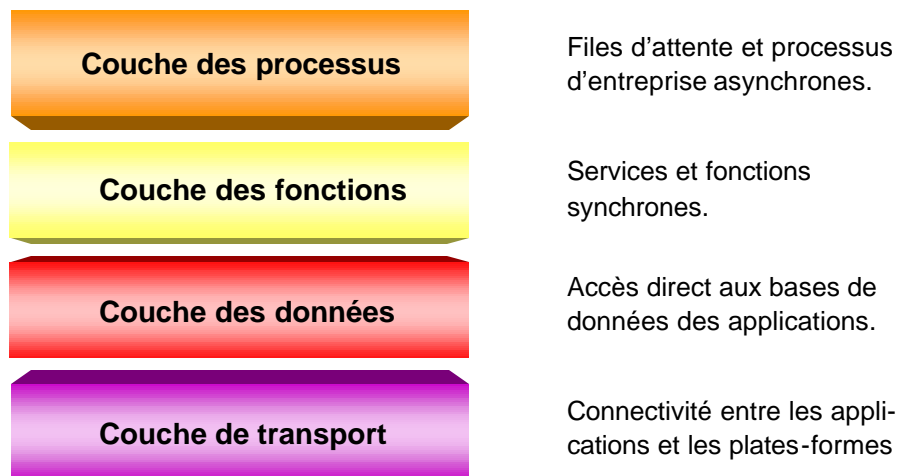
---

<sup>3</sup> « A General Purpose Registry/Repository Information Model »; Len Gallagher, Lisa Carnahan; octobre 2000; National Institute of Standards and Technology.

des systèmes hérités, mais il est bâti à partir de données extraites périodiquement des magasins de données opérationnelles qui ont été harmonisées en termes de format et de concept. L'entrepôt de données doit être entièrement compatible avec les métadonnées enregistrées dans le portail, pour que les utilisateurs puissent en avoir une vue cohérente. Pour les requêtes simples, cette transformation pourrait être effectuée dynamiquement, mais cela n'est habituellement pas faisable pour les applications plus sophistiquées qui établissent des sommaires et des rapports.

### III. INTÉGRATION DES APPLICATIONS

28. Une fois établis une approche commune pour la *gestion des identités* et un dépôt uniforme de métadonnées, nous pouvons porter notre attention sur les applications qui, au sein de l'organisation, seront intégrées et présentées aux utilisateurs du portail Web. Il s'agit souvent d'applications héritées, basées sur des technologies diversifiées et roulant sur de nombreuses plates-formes. En règle générale, ces applications ne sont pas conçues pour interagir avec le Web et il est fort probable qu'elles sont incompatibles entre elles pour ce qui est de la définition des entités et des processus d'entreprise.



#### *Couches d'intégration des applications*

29. L'intégration peut s'effectuer à plusieurs niveaux entre les applications, et nécessiter des efforts techniques importants. Le modèle ci-dessous présente quatre niveaux logiques, chacun pouvant nécessiter des solutions techniques différentes :

**Couche de transport :** C'est la couche de base qui sert à harmoniser les protocoles de communication et assure la connectivité entre les applications et les plates-formes. Outre les protocoles de base, comme la pile TCP/IP, cette couche peut comprendre des protocoles de courtage de messages synchrones (RPC, COM, CORBA) et une infrastructure de files d'attente de messages asynchrones (MQSeries, messagerie de courrier électronique).

**Couche des données :** Dans certains cas, il est plus approprié d'interagir directement avec les bases de données des applications, sans passer par les applications elles-mêmes. Comme les règles d'entreprise des applications sont contournées, il est souvent nécessaire de restreindre cet accès au mode « lecture seule », afin de préserver l'intégrité des bases de données. Cette couche utilise des protocoles standard comme le langage SQL et des interfaces communes comme ODBC.

**Couche des fonctions :** Dans un monde idéal, les applications offrent des interfaces « appelables » qui offrent des fonctions d'entreprise complètes au portail Web. Le mot « complètes » signifie que l'interaction a un contenu concret en termes

opérationnels (p. ex., ajout d'un client), et que toutes les règles et contraintes d'entreprise qui préservent l'intégrité de l'application sont appliquées. Ces fonctions travaillent habituellement en mode synchrone. En d'autres mots, les processus qui les appellent peuvent en déterminer l'échec ou la réussite.

Couche des processus : Cette couche s'occupe de l'ordonnancement des fonctions et des services, dans l'architecture. D'un point de vue hiérarchique, il s'agit d'une couche d'abstraction<sup>4</sup>, placée au-dessus des applications qu'on veut intégrer et notre article traite essentiellement de cette couche. Celle-ci est implémentée par la technologie EAI (*intégration des applications d'entreprise*) qui gère les flux et les messages au travers des applications, et qui dessert également les interactions du portail Web avec les utilisateurs.

30. Nous décrivons maintenant un scénario relativement simple s'appuyant sur ce modèle de base, qui permet d'intégrer des applications dans une architecture de portail Web. Cet exemple décrit les fonctions administratives requises pour ajouter un nouvel employé dans l'organisation et lui fournir les services requis pour l'exécution de son travail. Ces fonctions requièrent des interactions avec plusieurs applications situées dans différentes sections opérationnelles qui ne sont actuellement pas intégrées.

#### A. Exigences

31. Il s'agit d'offrir aux nouveaux employés un guichet unique quand ils se joignent à l'organisation, et de leur permettre de réaliser les processus administratifs habituels via les services fournis par l'intranet de l'entreprise.

32. Voici les différentes applications en cause :

- ?? Sécurité du bâtiment : Pour obtenir une carte d'identité avec photo, afin de pouvoir accéder au bâtiment, il faut vérifier le dossier de sécurité de la personne et émettre une carte d'identité. Ce processus est requis pour toutes les personnes qui doivent accéder régulièrement au bâtiment, y compris les employés, les contractuels et le personnel chargé de l'entretien du bâtiment.
- ?? Accès informatique : Les employés et les contractuels qui doivent accéder à l'infrastructure informatique reçoivent un code d'utilisateur. Ce code d'utilisateur leur donne accès au système de courriel, et peut-être à un compte Internet ou à d'autres privilèges.
- ?? Enregistrement des employés : Pour les employés seulement : il faut aviser les systèmes administratifs qui gèrent la rémunération et les avantages sociaux.
- ?? Enregistrement des services : L'employé peut avoir la possibilité de s'enregistrer à divers services, comme les privilèges d'accès à la bibliothèque, la formation, etc. De préférence, ces enregistrements secondaires seraient effectués directement par l'employé, une fois les privilèges d'accès ci-dessus établis.

33. Dans cet exemple, le système *sécurité du bâtiment* est à l'extérieur de l'architecture du portail Web, sauf que sa base de données est en lecture seule, via l'infrastructure de l'entreprise. Cette base de données constitue un exemple d'*intégration de données* à l'aide d'une interface ODBC. Seule l'information de base

---

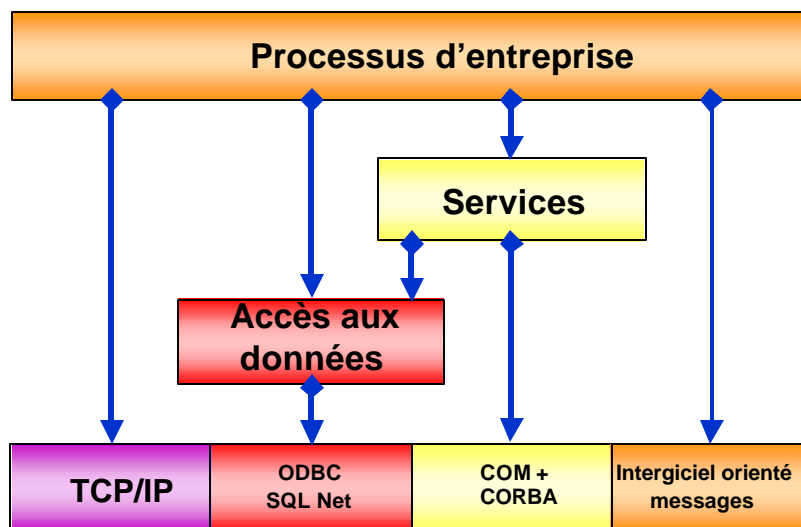
<sup>4</sup> En développement d'applications, on fait habituellement la distinction entre les phases compilation et exploitation. Les nouvelles pratiques de développement font maintenant la distinction entre la phase compilation (construction des composantes), la phase déploiement (assemblage dynamique) et la phase exploitation.

sur une personne est saisie pour l'émission d'une carte d'accès au bâtiment, par exemple l'information sur les contacts qui a été vérifiée manuellement sur un permis de conduire ou un passeport. Pour des motifs de confidentialité, aucun lien avec ces autres identificateurs n'est conservé, mais un code d'utilisateur unique est émis et codé sur la carte d'identité avec photographie.

34. Un administrateur qui a accès au portail Web peut effectuer le processus d'enregistrement de l'employé. Les processus d'entreprise sont habituellement conçus de telle sorte que les employés eux-mêmes doivent saisir et mettre à jour la majeure partie de l'information administrative requise. Toutefois, tant que l'employé n'a pas obtenu un code d'utilisateur lui permettant d'accéder au système informatique, il lui est évidemment impossible d'accéder à ces services.

35. L'administrateur utilise donc le service d'enregistrement du portail pour effectuer cette tâche. L'employé doit présenter sa carte d'accès au bâtiment (afin de bien s'identifier), puis son numéro d'identité est saisi à l'aide d'un lecteur de carte. Ce numéro est vérifié par rapport au contenu de la base de données de sécurité du bâtiment, après quoi le processus d'enregistrement est lancé. L'émission d'un code d'utilisateur du système informatique est un exemple de fonction pouvant être appelée directement par le portail Web en temps réel, via une connexion COM avec une fonction du système d'exploitation. Cet exemple illustre bien ce qu'on entend par *intégration fonctionnelle*. Le système d'exploitation comporte toutes les fonctionnalités requises pour émettre le code d'utilisateur, enregistrer le mot de passe initial et mettre à jour les répertoires sous-jacents dans l'infrastructure. La couche portail enregistre le code d'utilisateur dans sa base de données d'enregistrement, une fois ce code émis.

36. Le service d'enregistrement du portail doit également aviser le système de rémunération et d'avantages sociaux, afin de lui signaler l'ajout d'un nouvel employé, mais ce système est intégré moins étroitement au portail que ne l'est le processus d'enregistrement permettant d'accéder au système informatique. Dans ce cas-ci, on utilise un formulaire affiché à l'écran pour saisir l'information nécessaire (le formulaire peut déjà être en partie rempli automatiquement si l'information est déjà disponible pour le portail) et une transaction est effectuée, afin d'envoyer cette information au système de rémunération et d'avantages sociaux. Cette transaction est mise en file d'attente par le système EAI, puis est transmise de manière fiable lorsque le système de rémunération et des avantages sociaux est prêt à le recevoir. Dans les faits, ce système peut être une application héritée, fonctionnant par lots, qui est appelée périodiquement pour traiter toutes les transactions accumulées depuis un certain temps. Ce mécanisme de transmission asynchrone est un exemple d'*intégration des processus*, qui est réalisé à l'aide d'une solution intergicelle orientée messages, comme MQSeries.



*Interactions entre les composantes*

37. Le diagramme ci-dessus illustre schématiquement les différentes interactions entre les composantes. On y présente plusieurs technologies dans la couche de transport qui permettent aux composantes de communiquer entre elles, toutes ces technologies pouvant être appelées « *intergicielles* ». Ces diverses technologies peuvent être incorporées dans un même produit *EAI* qui pourrait comprendre des mécanismes permettant d'exécuter la transformation des données et l'établissement de règles d'entreprise.

## **B. Transformation des données**

38. Pour intégrer des applications qui ont été conçues à des moments différents et à des fins diverses, il faut inévitablement examiner en détail les concepts sous-jacents aux données et les formats d'implémentation. Le produit *EAI* devrait offrir des fonctionnalités poussées de transformation et de reformatage des données, car celles-ci peuvent devoir circuler entre différentes applications. C'est pourquoi une composante clé de l'architecture du portail Web est la *base de métadonnées intégrée* (IMDB). La base IMDB constitue le dépôt des règles de transformation des données, qui permettent la traduction des données, le transfert en direction et en provenance des applications existantes et l'harmonisation des concepts et des formats.

## **C. Règles d'entreprise**

39. La gestion des processus d'entreprise va bien au-delà de la simple transformation des données et de leur transfert aux applications appropriées, en vue de leur traitement. Même dans l'exemple très simple présenté ci-dessus, les identités doivent être vérifiées et les transactions ordonnancées, puis distribuées aux autres composantes. Cet ordonnancement peut dépendre du résultat de fonctions effectuées au préalable, ou de valeurs spécifiques entrées pendant une session interactive avec l'utilisateur.

40. De manière plus générale, la solution *EAI* complète peut comporter des réponses automatiques aux événements. Par exemple, des applications conçues pour faire le traitement différé des transactions peuvent être appelées à des moments précis de la journée, ou lorsqu'un certain volume de transactions a été accumulé. C'est le type le moins étroit de couplage entre composantes, mais cela évite les goulots dans le traitement qui peuvent ralentir le temps de réponse en ligne.

41. Ce mode d'intégration, piloté par les événements, se situe à l'extérieur de l'architecture du portail Web et est désigné par l'expression *gestion des flux de travail*. Les solutions *EAI* qui fonctionnent à ce niveau utilisent un modèle « *publication-abonnement* » pour déclencher des processus asynchrones ou acheminer des messages entre processus. Quand les messages sont sous forme de documents complets ou de formulaires d'entreprise, acheminés vers des utilisateurs dans le système, la technologie principale peut être alors l'infrastructure de courrier électronique.

42. L'architecture du portail Web comprend un dépôt dans lequel sont enregistrées les règles d'entreprise qui fonctionnent à un niveau plus granulaire, à l'intérieur de la *couche des processus*. En d'autres mots, ils indiquent la séquence des décisions et des actions qui se produisent à l'intérieur d'un processus d'entreprise.

## **D. Langage XML**

43. Bien que l'intégration des applications puisse se faire à l'aide de diverses composantes et technologies de messagerie, elle requiert des compétences poussées et donne lieu à des systèmes qui sont complexes et difficiles à maintenir. Dans le domaine des solutions *EAI*, on constate une tendance marquée vers la réduction de cette complexité : on a recours à des approches axées sur des spécifications et sur l'utilisation d'interfaces graphiques pour la conception et la construction des solutions.

44. Pour décrire cette architecture, nous avons défini plusieurs types différents de spécifications. Dans un cas concret, ces spécifications s'appuieraient sur la définition des interfaces de dialogue avec l'utilisateur, la

configuration des écrans, des états et des rapports, la structure des messages et le format des transactions, pour n'en nommer que quelques-uns. Pour simplifier cette prolifération de spécifications de métadonnées et de règles d'entreprise, le langage XML (*eXtensible Markup Language* – langage de balisage extensible) s'impose comme une syntaxe commune permettant d'exprimer ces définitions.

45. En plus d'exprimer les spécifications en langage XML, on peut utiliser cette même syntaxe pour les messages qui circulent entre les composantes et pour établir le contenu des données. En d'autres mots, le langage XML est en train de devenir la *lingua franca* pour toutes les couches dans le modèle d'intégration. Les avantages sont évidents : universalité, indépendance à l'égard des fournisseurs, communication plus facile entre les plates-formes et sur Internet.

46. Toutefois, l'utilisation d'une syntaxe normalisée représente seulement une dimension de la normalisation. L'utilisation du langage XML a donné lieu à l'émergence de plusieurs sous-langages qui portent sur des aspects différents du traitement des données, de leur transformation, de leur acheminement et de leur définition. De plus, d'importants efforts sont déployés afin d'introduire des définitions standard pour les documents ou les messages eux-mêmes, afin que l'intégration interorganisationnelle (B2B) puisse converger vers des transactions d'entreprise communes.

47. Un langage convenant à l'intégration des applications est le XLST (*XML Stylesheet Language Transformation* – transformation des langages de feuille de style XML), qui est une norme conçue par le World Wide Web Consortium (W3C), et qui recouvre à la fois les règles d'entreprise et la transformation des données. Le XLST est un langage conçu pour transformer un document XML en un autre, et permettre d'en modifier tant la structure que le contenu.

#### IV. APPLICATION DE L'ARCHITECTURE AUX SYSTÈMES STATISTIQUES

48. Les exemples ci-dessus sont de nature administrative, et on les retrouve dans les exigences opérationnelles de toute organisation. Dans le cas d'une organisation statistique, on peut également appliquer cette architecture quand on intègre l'infrastructure statistique, qui est composée de registres, du traitement des enquêtes, et des interactions avec les répondants. À maints égards, une infrastructure statistique est toutefois plus exigeante qu'un système administratif type. Dans cette section, nous décrivons quelques-unes des exigences additionnelles dont il faut tenir compte.

49. Par exemple, supposons qu'il y ait un portail commun pour les répondants individuels et les entreprises qui doivent s'inscrire pour fournir des données par voie électronique pour une ou plusieurs enquêtes réalisées par l'organisation statistique. Le portail n'offre aucune fonction de traitement statistique proprement dite, mais on peut le concevoir pour qu'il gère les fonctions communes de *gestion des contacts*, de *gestion des réponses* et de *gestion initiale des données*, ces fonctions desservant les systèmes de traitement en arrière-plan.

50. Par exemple, le portail pourrait gérer les registres suivants :

- *Registre des contacts.* Ce registre contient de l'information sur toutes les personnes qui représentent les répondants (entreprise ou ménage) et qui sont autorisées à fournir des données dans le cadre d'une enquête particulière. En règle générale, il peut y avoir plusieurs contacts pour une enquête donnée, notamment dans le cadre des enquêtes auprès des entreprises, et un même contact peut être le répondant pour plusieurs enquêtes. Les données enregistrées dans le portail sur chaque contact ne sont pas confidentielles (au sens statistique) et visent uniquement à authentifier et identifier les répondants. Par exemple, ces données pourraient comprendre le *nom*, l'*adresse électronique* et le *numéro de téléphone*.

- *Registre des enquêtes.* Chaque instance d'une enquête (cycle) qui accepte des transactions électroniques est enregistrée ici. Le registre pourrait également identifier les définitions XML qui sont valides pour cette enquête. En d'autres mots, le registre indique les formulaires électroniques qui représentent le questionnaire et ses métadonnées associées. En outre, il indiquerait les applications d'arrière-plan responsables du traitement des réponses et des autres informations publiques qui pourraient être utiles pour les contacts des répondants (documents d'information standard sur l'enquête, déclarations relatives à l'utilisation des données, politique de respect de la vie privée, etc.).
- *Registre des réponses.* Ce registre fait le suivi des instances de réponse, et il maintient un journal des interactions avec un contact spécifique pour une instance d'enquête spécifique. Cette information est conservée aux fins de contrôle des enquêtes, par exemple pour appliquer les règles d'entreprise concernant les réponses multiples ou pour identifier une mesure devant faire l'objet d'un suivi téléphonique.

51. Il s'agit là d'une description très sommaire d'un portail pour les enquêtes, qui pourrait gérer l'enregistrement des *contacts*, grâce à un gestionnaire (ou système automatisé) d'enquêtes permettant aux contacts de s'identifier quand ils répondent électroniquement à une enquête, et qui pourrait gérer les sessions pour le contact pendant le cycle de réponse. Dans une simple enquête par courriel, le portail utiliserait le registre des contacts pour envoyer les questionnaires électroniques aux répondants (avec courriel sécurisé si le questionnaire contient des données sur les répondants). Par exemple, le questionnaire peut comprendre le *numéro d'entreprise*, ainsi que l'*identificateur d'instance d'enquête*. Ces identificateurs permettent au contact de s'enregistrer et de soumettre ses réponses par l'intermédiaire du portail. Les réponses seraient transmises par formulaire électronique chiffré, acheminé directement à un système de traitement en arrière-plan.

52. Dans cet exemple, le portail ignore tout du contenu de l'enquête et s'occupe uniquement de l'information de contrôle de l'enquête. L'enquête même serait gérée par un outil commercial de formulaires électroniques, sécurisés par un logiciel de chiffrement. Toutefois, si on utilise cette architecture de portail pour gérer le contenu des données dans les réponses, il y a lieu de tenir compte des points suivants.

#### **A. Harmonisation<sup>5</sup> des concepts**

53. Si on veut offrir un portail commun pour plusieurs enquêtes (p. ex., enquêtes auprès des entreprises), il faudra accorder une attention accrue à l'harmonisation des données. Comme nous l'avons souligné précédemment, les outils offerts par le portail peuvent comprendre des services de base de restructuration et de reformatage des transactions, afin que celles-ci puissent être interfacées avec les systèmes de traitement existants, mais l'utilisation de concepts différents serait une source de confusion pour les répondants.

54. On devrait chercher en premier lieu à harmoniser les informations d'identification et de classification qui sont habituellement enregistrées dans les registres statistiques, comme le *registre des entreprises*, le *registre des adresses* et le *registre des données fiscales*. Ces registres contiennent des données administratives qui ont été fournies à l'origine par le client (souvent par l'intermédiaire de canaux ou de sources multiples), ou assemblées et interpolées par le bureau statistique.

---

<sup>5</sup> Par harmonisation, on entend le processus intellectuel humain qui produit des règles de regroupement cohérent et uniforme. Ces règles de regroupement, ou d'assemblage, peuvent, une fois définies, être appliquées à la construction d'entreprises à partir d'entités juridiques et d'entités opérationnelles identifiées au préalable. L'harmonisation peut également être appliquée à des problèmes plus abstraits, notamment la conception et l'assemblage de systèmes de classification à l'aide de classes de composantes.

55. Dans de nombreux pays, la législation sur la protection de la vie privée exige que les dossiers administratifs puissent, une fois sécurisés adéquatement, être accessibles aux personnes visées par ces dossiers, afin qu'elles puissent corriger toutes données erronées. Bien que les données statistiques soient souvent soustraites à l'application de ce type de loi, cette exemption ne couvre pas nécessairement les données obtenues par l'intermédiaire de sources administratives.

56. La normalisation des concepts et des classifications de données dans les dossiers administratifs sort du domaine couvert par cet article, mais il convient néanmoins de noter qu'on utilise de plus en plus le XML comme syntaxe pour la description des interfaces publiques des données. Par exemple, le langage XBRL (*eXtensible Business Reporting Language*) est une norme en cours de développement, et permet de définir les plans comptables et la structure des états financiers et des bilans des entreprises. Si ces normes font l'objet d'un large consensus et deviennent d'un emploi généralisé, on pourrait y recourir pour harmoniser les concepts, non seulement dans les organisations statistiques, mais dans de nombreux ministères et organisations gouvernementales.

## **B. Synchronisation**

57. L'intégration des applications statistiques, notamment celles faisant appel aux registres administratifs, doit permettre la résolution des problèmes associés à l'utilisation d'informations recueillies à divers moments. En ce qui concerne l'architecture du portail Web, on devrait adopter une politique de *temps courant*. En d'autres mots, le portail devrait présenter seulement les données administratives les plus récentes (ou actualisées) aux utilisateurs externes, et indiquer de manière appropriée la *période de référence* si les données ne sont pas actuelles ou s'appliquent à une plage de temps précise.

58. En règle générale, la rationalisation des données statistiques pour différentes périodes de temps requiert des processus complexes d'ajustement temporel et saisonnier. Bien que de telles méthodes permettent d'améliorer la qualité des estimations et de l'imputation des données statistiques, les valeurs obtenues ne devraient normalement pas être exposées aux répondants.

59. Par conséquent, on se doit d'établir une distinction nette entre l'information contenue dans les registres du portail Web et les bases de données de référence usuelles (qui sont également souvent appelées des registres), dans les organisations statistiques. Les bases de données de référence, comme le *registre des entreprises*, contiennent des enregistrements historiques de la structure des entreprises, structure qui varie avec le temps. Dans le même ordre d'idées, les bases de données de référence géographique contiennent des définitions (variant dans le temps) des frontières et des zones administratives. Ce type de bases de données recoupe des concepts complexes qui sont intrinsèquement difficiles à intégrer et à synchroniser, et qu'on ne peut donc exposer directement aux répondants. Les registres du portail Web qui sont utilisés pour communiquer avec les répondants devraient plutôt être confinés à l'information que le répondant peut changer de manière dynamique, à mesure que les événements prennent place dans le monde réel.

## **C. Métainformation et métadonnées**

60. Les métadonnées statistiques varient également dans le temps (par exemple, classification des industries, des biens ou des professions), et il faut alors procéder au recodage ou à des exercices de concordance pour harmoniser les données, ce qui peut s'avérer fort complexe. Ces complexités s'insèrent mal dans un portail Web en temps réel qu'on voudrait utiliser pour présenter une vue cohérente aux répondants.

61. Par conséquent, les métadonnées enregistrées dans l'architecture du portail doivent être confinées aux définitions courantes utilisées dans les interfaces de dialogue avec l'utilisateur, plutôt que comprendre l'ensemble des métadonnées requises par l'organisation statistique.

## V. CONCLUSION

62. Comme nous le décrivons dans cet article, on peut en général recourir à l'architecture d'un portail Web pour offrir une interface avec les applications existantes, pourvu qu'on observe les principes de conception suivants :

- *Être axé sur l'utilisateur* : Aux yeux des utilisateurs, le portail doit constituer un système cohérent et unique. On doit donc déployer des efforts importants pour définir des concepts de données communs et comprendre la transformation de ces concepts, quand ils sont interfacés avec les applications existantes.
- *Simplicité* : La logique d'entreprise intégrée dans le portail doit être confinée aux interactions avec les utilisateurs et au contrôle ou à l'ordonnancement des fonctions. Les règles ou processus d'entreprise complexes, qu'on retrouve dans les applications sous-jacentes, ne devraient pas être reproduits ou exposés dans le portail.
- *Actualité* : Un portail Web est une interface en temps réel, pour l'ensemble des utilisateurs. Par conséquent, il devrait présenter uniquement l'information actualisée et des concepts simples de temps.
- *Utilité* : L'interface d'un portail Web devrait présenter un ensemble clair de services à ses utilisateurs. Le type de portail décrit dans cet article permet aux utilisateurs d'effectuer des tâches. Des services comme l'*enregistrement*, le *remplissage de formulaire*, leur *présentation* et leur *vérification* peuvent être généralisés à un large ensemble d'applications et contribuer à assurer une interface uniforme pour les utilisateurs.

63. Nous avons montré que les normes nouvelles, comme le langage XML et le recours aux technologies commerciales, jouent un rôle majeur dans le développement des systèmes de ce type.

64. Nous espérons que cette architecture constituera une piste de travail pour Statistique Canada, et lui permettra d'expérimenter les technologies Web tout en préservant ses investissements dans les applications existantes.