

E-GOUVERNEMENT EN BELGIQUE

Un retour riche d'expériences

La sécurité sociale de Belgique a mis en place un système de méta-information multilingue pour la documentation technique des bases de données. Le système sert aussi lors de l'échange de messages électroniques avec les assurés sociaux internautes.

Avec l'émergence et la complexité croissante des applications informatiques, la documentation des programmes est plus que jamais vitale, quel que soit le secteur d'activité concerné. Nous présentons ici les enseignements méthodologiques de la mise en place d'un système de méta-information sur le portail de la sécurité sociale belge. Ce système a pour objet la documentation technique des bases de données de la sécurité sociale et des messages électroniques correspondants. L'information échangée entre les citoyens et l'administration permet le prélèvement des cotisations sociales ainsi que leur redistribution aux assurés sociaux, ce qui implique le traitement annuel d'environ 38 milliards d'Euros. Les enjeux sont donc considérables, l'information administrative étant créatrice de droits et de devoirs(2).

Les systèmes de méta-information

La "méta-information" est souvent entendue comme "information sur l'information". Nous retenons ici la définition suivante : un système de méta-information est un système documentaire automatisé destiné à décrire un ensemble d'informations et ce faisant, à l'interpréter en vue d'en faciliter la gestion. Le recours à de tels systèmes est stratégique lorsque l'information est un instrument d'action sur le réel(3).

La conception d'un système de méta-information s'inscrit dans le cadre d'une stratégie de gestion. Les coûts correspondants émanent des opérations d'analyse, de conception et de



Isabelle Boydens

Chargée de cours à l'Université Libre de Bruxelles et consultante (1)

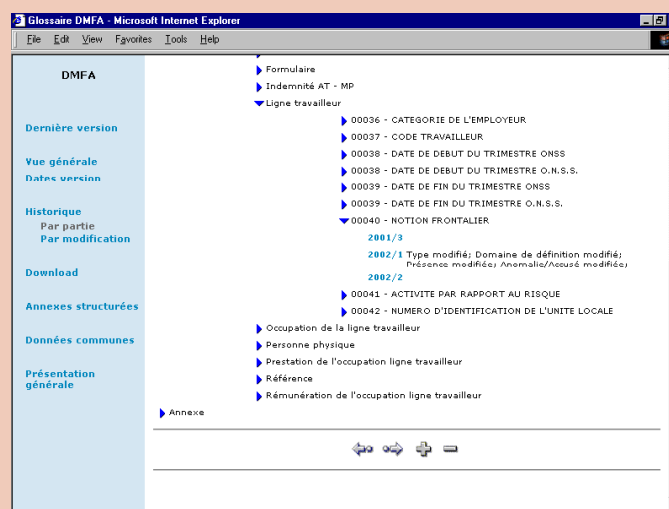
“

La conception d'un système de méta-information s'inscrit dans le cadre d'une stratégie de gestion

”

1/ Isabelle Boydens est consultante à la section "recherche" de la SmalS-MvM, société informatique prestataire de services pour la sécurité sociale belge. 2/ Boydens I., Informatique, normes et temps. Bruxelles : Bruylant, 1999. 3/ Ainsi, pendant la guerre du Golfe de 1991, environ 28.000 des 40.000 containers d'armements militaires américains envoyés au Moyen-Orient furent inspectés et inventoriés manuellement tant l'interrogation des bases de données correspondantes, censées en répertorier le contenu, donnait lieu à des résultats incohérents. D'où cette remarque amère : "in general, the physical movement of material is faster than the movement of the supporting information." Madnick S. E., "The Voice of the Customer : Innovative and Useful Research Directions". In Agrawal R., Baker S., et Bell D., eds, Proceedings of the 19th Conference on Very Large Databases, Dublin, VLDB, 1993, p. 702. Il va de soi que si chaque base de données avait été correctement documentée et interprétée, l'exploitation automatisée de l'information aurait été plus aisée.

Figure 1 • Modifications apportées aux données version par version ("delta")



“
L'application s'adresse
à la fois aux informati-
ciens et aux citoyens

”

maintenance. Les bénéficiaires escomptés tiennent à une meilleure interprétation de l'information, à une réutilisation plus aisée des applications pré-existantes, à une crédibilité accrue du système et à une diminution des coûts de gestion (corrections a posteriori de la base de données, réparation des préjudices dus à la diffusion de données incorrectes...).

“
Il s'agit de modéliser la
connaissance et les
processus qui l'engen-
drent

”

A l'heure actuelle, dans la mouvance du "web sémantique", les normes en la matière sont devenues nombreuses(4). Les unes offrent des syntaxes génériques permettant le déploiement de méta-données, tels que RDF (Ressource Description Framework), recommandation du World Wide Web Consortium (voir : <http://www.w3.org/RDF/Group/WD-rdf-syntax/>) ou encore les "topic maps", issus de l'organisation ISO. Les autres sont ciblées sur des domaines d'application particuliers, tels que le Dublin Core (voir : <http://dublincore.org/>), l'e-Gouvernement Metadata Standard (e-GMS) ou encore, le Digital Imaging Metadata Standard (DIG35). La plupart

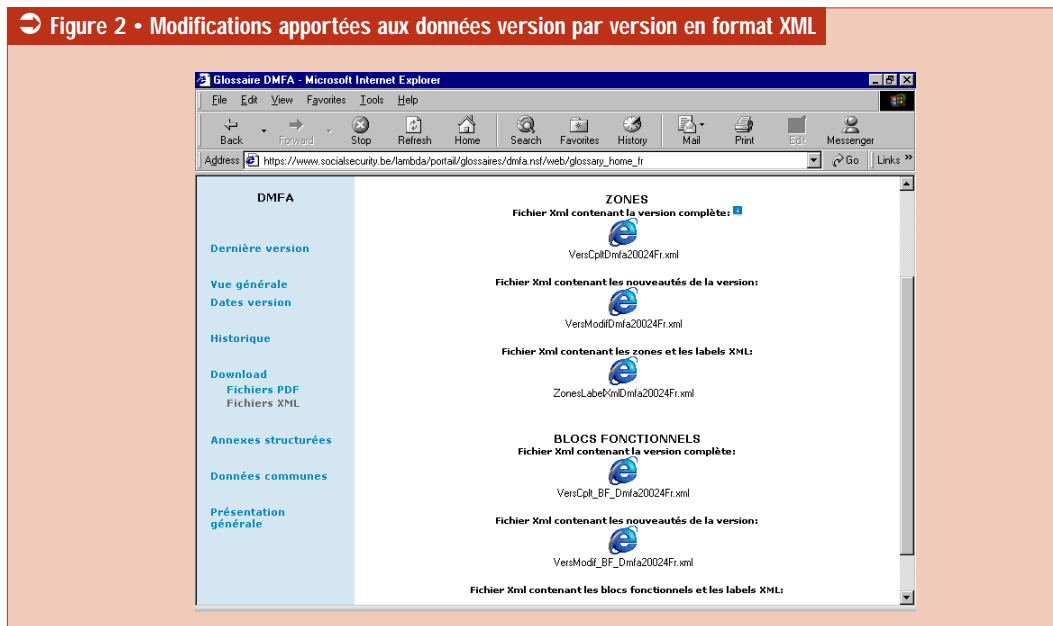
de ces normes ont recours au langage XML (eXtensible Markup Language).

Retour d'expérience

L'application présentée ici s'adresse à la fois aux informaticiens en charge de la gestion des bases de données administratives et, via le portail de la sécurité sociale, aux citoyens en charge de l'envoi des déclarations électroniques à l'administration. L'objectif était de faire en sorte que tous travaillent sur une base commune. Ce système documentaire s'apparente à un dictionnaire électronique de données. Il vise à : automatiser partiellement les procédures de saisie, de traduction et de validation de la documentation ; à en renforcer l'intégrité(5) ; et à en gérer les versions au fil des modifications législatives. Il s'agit de modéliser la connaissance et les processus qui l'engendrent. Ainsi, le dictionnaire inclut à la fois des informations descriptives (par exemple, le domaine de définition d'un champ) et fonctionnelles (par exemple, la spécification formelle des contrôles destinés à tester les déclarations entrantes). Par

4/ Abrams C., Metadata Management : Crucial for the Real-Time Enterprise. Business Intelligence 2003. Amsterdam : Gartner, 3-4 février 2003. 5/ Boydens I., Pirotte A. et Zimanyi E., Managing Constraints Violations in Administrative Information Systems. In Proceedings of the 7th IFIP 2.6 Conference on Data Semantics, DS-7, Leysin, October 1997. Chapman & Hall, 1997, p. 241-264. 6/ L'application a été développée au sein de l'Unité d'exploitation " Développement des applications et consultation " de la SmIS-MvM. Le lien suivant donne accès à une application de ce type en ce qui concerne la base de données DmfA (déclaration multifonctionnelle) : https://www.socialsecurity.be/lambda/portail/glossaires/dmfa.nsf/web/glossary_home_fr

Figure 2 • Modifications apportées aux données version par version en format XML



“
La gestion des versions
est fondamentale

”

ailleurs, les schémas des messages XML échangés entre les citoyens et l'administration sont générés à partir du dictionnaire électronique. Celui-ci est accessible en consultation (partie "front end") sur le portail de la sécurité sociale belge (<http://www.socialsecurity.be>)⁽⁶⁾ et a été développé via un système de "groupware-workflow"⁽⁷⁾.

Nous présentons successivement les fonctionnalités suivantes : gestion de champs multilingues, gestion des versions, mise en place de mécanismes d'héritage, application du concept de WOPM (Write Once Publish Many), système de recherche multibase et enfin, déploiement d'un workflow de validation documentaire.

La gestion de champs multilingues

En Belgique, la documentation technique doit être diffusée dans les différentes langues nationales. Des fonctions multilingues sont par ailleurs requises dans tout contexte supranational. Des tables contrôlées multilingues permettent, lors de la saisie des définitions, d'intégrer l'informa-

tion dans une langue et d'obtenir ses contreparties dans les autres langues. Ceci permet de minimiser la charge de travail manuel, d'accélérer le processus de saisie et de renforcer la cohérence de l'ensemble.

La gestion des versions

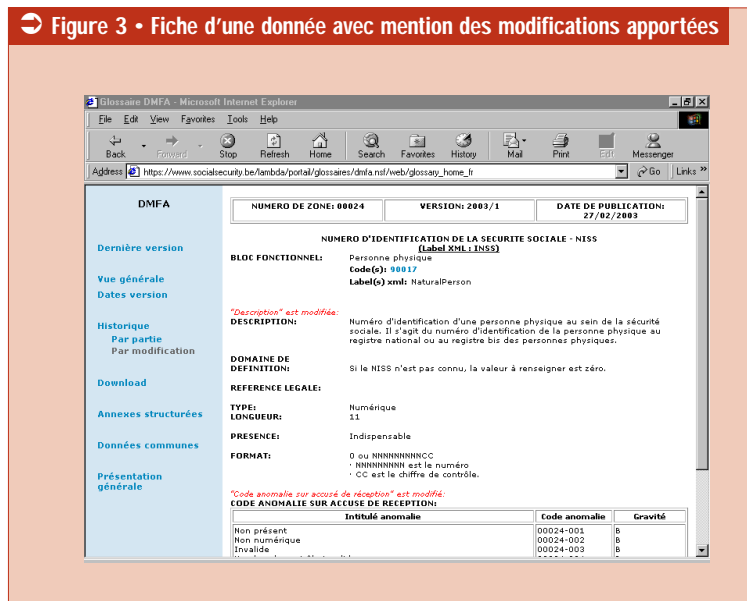
La gestion des versions est fondamentale dans le domaine administratif⁽⁸⁾. En effet, la législation évolue fréquemment et toutes ses versions successives doivent être conservées au moins durant la période de prescription (par exemple, lorsqu'il s'agit de traiter des arriérés, il est fondamental de pouvoir retrouver les définitions antérieures de la base de données, les déclarations enregistrées ayant un statut légal de "force probante", c'est-à-dire qu'elles peuvent faire office de "preuve" lors d'un litige devant les tribunaux). Il est dès lors crucial d'identifier précisément les modifications apportées à chaque nouvelle version par rapport à la précédente (figure 1). Ce "delta" est par ailleurs diffusé sous format XML, de telle sorte que les modifications puissent être intégrées de façon semi-structurée dans les applicatifs encadrant les bases de don-

“
Il est crucial d'identifier
précisément les modi-
fications apportées à
chaque nouvelle ver-
sion

”

7/ Boydens I., Déploiement coopératif d'un dictionnaire électronique de données administratives, Revue Document Numérique (Numéro spécial : "Création et gestion coopératives de documents numériques d'information et de communication"). Paris : Editions Hermès Sciences, vol. 5, n° 3-4/2001, p. 27-43 (<http://dn.e-revues.com/>). 8/ Boydens I., Les bases de données sont-elles solubles dans le temps ? La Recherche hors série ("Ordre et désordre"). Hors série n° 9, novembre-décembre 2002, p. 32-34.

Figure 3 • Fiche d'une donnée avec mention des modifications apportées



“
Le dictionnaire électronique est destiné à documenter plusieurs dizaines de bases de données administratives

”

nées (figure 2). Chaque item spécifiant la description d'une donnée pour une version considérée renvoie à la fiche correspondante (dans la langue choisie par l'utilisateur) avec la mention des champs modifiés par rapport à la version immédiatement antérieure (figure 3).

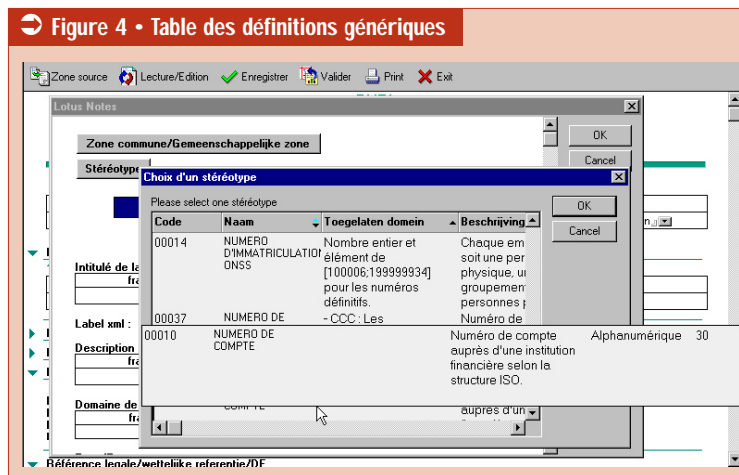
Héritage et réutilisation

Le dictionnaire électronique est destiné à documenter plusieurs dizaines de bases de données administratives répertoriant de nombreux champs communs dont certaines caractéristiques sont identiques (format, par exemple) et d'autres distinctes (caractère obligatoire ou facultatif d'un champ, par exemple). Un mécanisme d'héritage a dès lors été mis en place.

“
Le système permet d'assurer que les données communes reçoivent les mêmes valeurs

”

Figure 4 • Table des définitions génériques



L'héritage se définit comme la relation entre une classe A générique (que nous appelons ici "stéréotype") et l'ensemble de ses instances {a1, a2, ... an}, où les propriétés (p1, p2, ...pk) de la classe A constituent un sous-ensemble des propriétés de chaque objet instancié à partir de la classe A. Lors de l'instantiation, ce sous-ensemble de propriétés génériques peut être complété par un autre sous-ensemble de propriétés spécifiques à chaque instance

(p1+pa1, p2+pa2, ...pk+pan). Ce mécanisme est applicable à un nombre arbitraire de niveaux "méta".

Dans notre exemple, les valeurs des propriétés génériques ("nom", "domaine de définition", "description", "type", "longueur") du stéréotype "numéro de compte" sont stockées dans une table "contrôlée" de données structurées génériques (figure 4).

Ces propriétés génériques multilingues peuvent être héritées dans une fiche documentaire et complétées lors de l'instantiation par des propriétés spécifiques (dans l'exemple de la figure 5, en français : "intitulé" : "numéro de compte + de l'employeur" ; et sur la figure 6 "numéro de compte + de l'assuré social (victime)").

Les valeurs génériques et spécifiques sont ainsi concaténées en un champ semi-structuré. Ces fonctionnalités apportent des avantages en terme de temps de mise à jour (chaque valeur générique ne devant être encodée qu'une seule fois) et en terme de cohérence. Le sys-

tème permet d'assurer que les données communes reçoivent les mêmes valeurs et d'éviter les erreurs humaines inhérentes à l'encodage manuel.

Write Once Publish Many

L'application inclut des listes structurées (codes postaux, catégories d'activité, ...) qui, dans la pratique, doivent être diffusées à des fins documentaires (dans l'esprit d'un "système de méta-information") mais aussi en vue de tester les déclarations envoyées par les citoyens et stockées dans les bases de données. Afin de rencontrer les deux fonctions, l'application a été conçue dans l'optique du concept WOPM ("Write Once Publish Many") de façon à générer automatiquement une même table structurée (liste de codes postaux, par exemple) sous différents formats : formats ascii, XML, word, excel et PDF. La même source peut ainsi être utilisée au sein d'applications interdépendantes (figure 7).

Système de recherche "multibase"

Un outil de recherche "multibase" a été mis en place, permettant une recherche "full text" à travers le système documentaire intégré sur base de paramètres spécifiques (figure 8).

Workflow de validation

En raison des enjeux légaux, sociaux et financiers correspondants, chaque nouvelle version doit être validée par les informaticiens et les juristes concernés par celle-ci. En vue de struc-

Figure 5 • Exemple d'héritage d'une valeur générique (1)

turer cette validation, un système de workflow guide le déploiement du dictionnaire électronique (figure 9). Celui-ci s'inscrit dans le cadre d'un planning annuel de mise à jour, spécifiant de façon rigoureuse les périodes de mise à jour, de validation, de mise en acceptation et de mise en production. Le workflow est "pilote" de façon centralisée par une équipe dédiée à cette tâche et se déploie sur un mode décentralisé dans le cadre de l'extranet de la sécurité sociale. Lors de la création de chaque nouvelle version, l'historique des échanges entre les différents responsables est conservé, de façon à garder un suivi du processus d'interprétation. Une vue permet aux gestionnaires de suivre le nombre de "fiats"⁽⁹⁾ requis pour la publication d'une nouvelle version sur le portail de la sécurité sociale⁽¹⁰⁾.

Recommandations méthodologiques

Les systèmes de méta-information comportent potentiellement trois écueils. Le premier est lié

9/ Les "fiats" correspondent aux "accords" ou aux "validations" requises par les responsables des documents lors de la publication d'une nouvelle version (les juristes vérifient l'adéquation à la loi et les informaticiens vérifient l'adéquation formelle aux programmes). Pour chaque ensemble documentaire à mettre à jour, un ensemble de responsables en charge de la validation des modifications de contenu est désigné. Une vue permet aux gestionnaires de suivre le nombre de "validations" requises pour la publication d'une nouvelle version sur le portail de la sécurité sociale. 10/ Des procédures de décision sont prévues en cas de "non réponse" de l'un des acteurs dans les délais fixés ; par ailleurs une certaine flexibilité permet l'insertion d'un addendum entre deux versions officielles.

“
Un outil de recherche
"multibase" a été mis
en place

”

“
Le workflow est "pilote"
de façon centralisée
par une équipe dédiée

”

“
Les systèmes de méta-
information comportent potentiellement
trois écueils

”

au fait que ces systèmes sont extensibles à l'infini, surtout lorsque les champs à compléter sont "libres", le langage naturel étant son propre méta-langage. Ceci implique des coûts importants en termes de gestion, lorsque les mises à jour manuelles sont nombreuses. Le second écueil tient à ce que les méta-données peuvent être elles-mêmes erronées et incertaines. Lorsqu'elles sont d'ordre contextuel, leur validation ne peut faire l'objet de contraintes d'intégrité rigoureuses. Le troisième

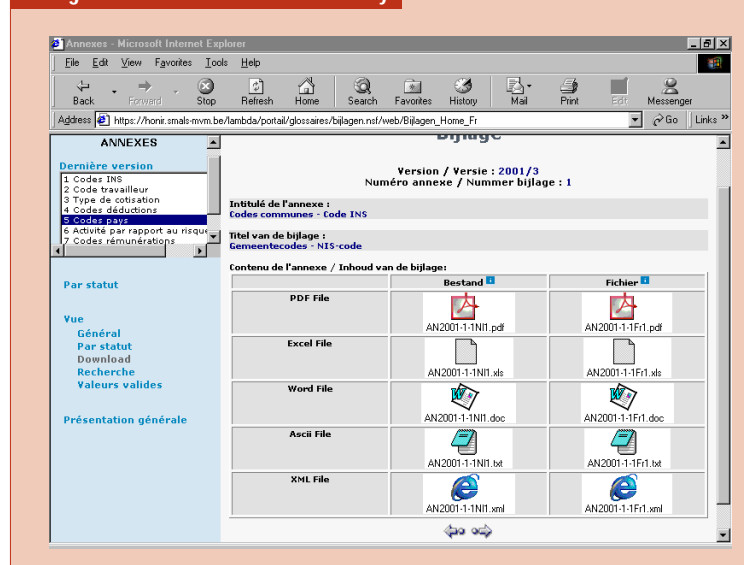
écueil tient au décalage temporel entre la mise à jour d'une donnée et de la méta-donnée correspondante, cette dernière, surtout lorsqu'elle se présente sous une forme textuelle, n'étant généralement créée qu'au terme d'une phase d'analyse.

Ainsi, dans une communication retentissante, "The Metadata Myth..."⁽¹¹⁾, plusieurs auteurs

“
Plusieurs auteurs évoquent les inextricables difficultés pratiques que soulève "l'usage abusif" des méta-données

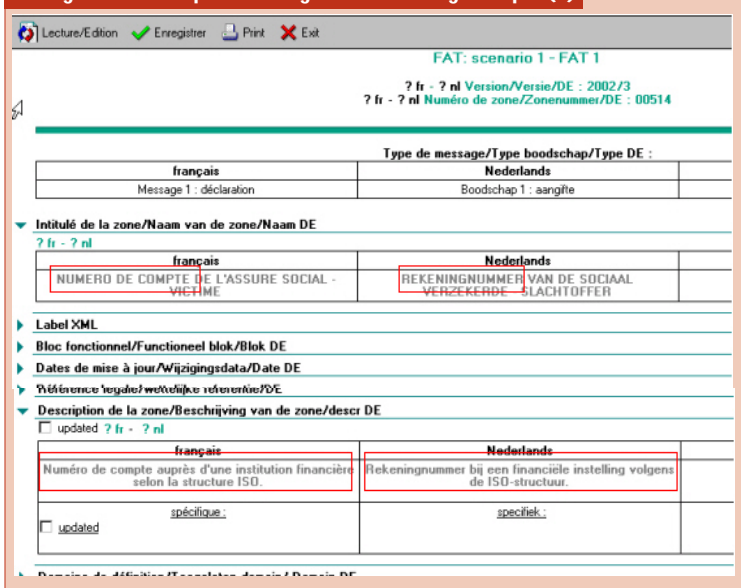
”

Figure 7 • Write Once Publish Many



11/ Foreman T.W., Wiggins H.V., Porter D.L., Metadata Myth : Misunderstanding the Implications of Federal Metadata Standards. Proceedings of the First IEEE Metadata Conference. Maryland : IEEE, 1996 (http://www.llnl.gov/liv_comp/metadata/ieee-md-4-96.html).

Figure 6 • Exemple d'héritage d'une valeur générique (2)



évoquent les inextricables difficultés pratiques que soulève "l'usage abusif" des méta-données. Dans le domaine des bases de données géospatiales exploitées par le Bureau of census et la National Aeronautics and Space Administration (NASA), la mise en place d'un système fédéral de méta-information pour lequel chaque nouvel enregistrement nécessitait l'intégration d'environ 300 méta-données a entraîné les avatars suivants :

coûts exorbitants en personnel et en ressources, lourdeur des mises à jour, ésothérisme de la documentation et, finalement, réduction considérable de l'échange des données. Cependant, la NASA n'a pas abandonné ce système qui a toutefois fait l'objet d'une simplification et d'une restructuration.

Sur la base des expériences en la matière, nous proposons les cinq recommandations suivantes :

Figure 8 • Ecran de recherche multibase

Possibilité d'étendre ou de restreindre la recherche (par défaut porte sur tous les dictionnaires)

Possibilité de préciser :

- dans quel type de document la recherche doit se faire (zones, annexes)
- dans quelle version (dernière version ou toutes les versions)
- dans quel(s) champ(s) on souhaite lancer la recherche en fonction du type de document
- les paramètres d'affichage des résultats
- la sortie souhaitée (interface Web → bouton "search" ou export en excel)

Search

GLOSSARIES
Select glossaries : ACR, DMFA, dsfat1 (Empty means Any Glossary)

OPTIONS
Document type :
Search only in last version : Yes

SEARCH
Search for the following word :
In fields : (Empty means Any Field)

RESULTS
Display for each doc the fields : Initulé de la donnée, Numéro de zone - Zonennummer

Search Export to excel Reset

“
Les méta-informations
générées automati-
quement sont moins
"coûteuses"

”

1. l'identification d'un ensemble minimal de méta-données obligatoires(12) ;
2. une préférence pour les méta-informations générées automatiquement (ou sur base de listes de valeurs contrôlées par exemple) car ces informations sont moins "coûteuses" en termes de mise à jour et plus fiables(13) ;
3. la création de plusieurs niveaux de méta-données adaptés en fonction des usages (méta-données

- génériques et spécifiques)(14) ;
4. une prise en compte continue de l'organisation dans lequel s'insère le système de méta-information, via un système de workflow de validation ;
5. la mise en place de liens directs entre les applicatifs documentés et les méta-données correspondantes (dans l'exemple présenté ici, les schémas des messages XML échangés entre l'administration et les citoyens sont générés à partir du dictionnaire de données).

Figure 9 • Workflow de validation

Instructions DmFA - DRS - Microsoft Internet Explorer

Validation Workflow Instructions & Glossaries

New Document Previous Set of Documents Next Set of Documents

By version	Date	Version	Sector	Subject
By date	20/06/2002	2002/2		
By author			DMFA	
By type				2 Validation Administrative Instructies Dmfa versie 2002/2 (Hans Verdonck 20/06/2002)
By receiver				Response to Validatie Administrative Instructies Dmfa versie 2002/2 (Stephane Carcan 20/06/2002)
# Fiats				Response to Validatie Administrative Instructies Dmfa versie 2002/2 (Stephane Carcan 20/06/2002)
	14/06/2002	2002/2	DMFA	
				1 Validation Dmfa Glossarium 2002/2 (Hans Verdonck 14/06/2002)
				Response to Validation Dmfa Glossarium 2002/2 (Dirk Wulpeutte 21/06/2002)
	07/06/2002	2002/2		
			RVA-DNEM	

Au-delà de l'application présentée dans cet article, ces recommandations s'appliquent à toute base de données empiriques dont l'interprétation est stratégique, en tant qu'instrument d'action sur le réel.

“
Ces recommandations
s'appliquent à toute
base de données
empiriques dont l'inter-
prétation est straté-
gique

”

Isabelle Boydens

Revue d'auteurs,
l'Informatique Professionnelle accueille
des opinions qui n'engagent pas
la rédaction.

12/ Ainsi le dictionnaire présenté ici documente environ 30 bases de données, incluant plusieurs centaines de champs et 14 zones (ou groupes de zones) descriptives obligatoires 13/ Dans l'application présentée, les valeurs de 5 groupes de zones sont mises à jour de façon semi-automatique (dates de modification, mise à jour sur base de listes fermées de champs contrôlés...). Par ailleurs, lors de la création d'une nouvelle version, toutes les valeurs de la version antérieure sont héritées de telle sorte que seules les modifications à apporter doivent être saisies ou générées. Enfin, à l'intérieur d'une même version, le mécanisme d'héritage à partir de concepts génériques représente également une économie sur le plan de la mise à jour et contribue par ailleurs à assurer l'intégrité de l'ensemble (le contenu d'environ 2000 zones est généré via ce mécanisme d'héritage). 14/ Dans l'application présentée, 4 groupes de zones sont à usage interne (les autres étant publiques sur le portail de la sécurité sociale) et définissent les paramètres permettant la génération des schémas XML ainsi que certains contrôles internes appliqués aux bases de données.