



Blog Talks Publications ∨ Tools ∨ Radar / Plan ∨ Team About

De la production industrielle à la production d'information : analogies, paradoxes et enseignements opérationnels

Posted on 2013-07-31 by [Isabelle Boydens](#)



Les normes internationales actuelles en matière de "data quality" puisent leur origine au début du XXème siècle, dans la foulée de la révolution industrielle et du taylorisme, alors que la production et le montage en grande série réclament l'interchangeabilité des pièces. Aussi a-t-on recours au travail "à tolérances", dans le cadre duquel les dimensions des pièces usinées peuvent varier dans des limites définies et normalisées. L'objectif consiste à atteindre un seuil de précision compatible avec l'impératif d'interchangeabilité, tout en minimisant les coûts correspondants.

Dès l'origine sont donc apparus quelques grands principes fondamentaux en matière de qualité des bases de données : l'idée de "**one best**" ("meilleur relativement") indique que la **perfection est une non valeur**". En effet, le concept de norme est intrinsèquement historique et évolutif ([Boydens, I., L'océan des données et le canal des normes. In Annales des Mines, Paris, 2012](#)). On se trouve toujours face à

un **arbitrage de type "coûts-bénéfices"**, eu égard à une demande perçue, de là, le concept de **"fitness for use"** ("adéquation aux usages").

Une comparaison entre production industrielle et production d'information permet de dégager, à la lumière de normes et d'exemples pratiques récents, plusieurs analogies et paradoxes dont découlent quelques **"bonnes pratiques" opérationnelles** en matière de gestion de la qualité des bases de données.

- **Sur le plan financier et comptable, contrairement à un produit matériel, celui qui vend une information ne s'en dépossède pas... mais, au contraire, peut avoir intérêt à la "partager"**

Alors que celui qui vend un produit matériel s'en dépossède, celui qui transmet une collection de données (laquelle peut être exploitée à des fins diverses par plusieurs utilisateurs simultanément, en mode "partagé") ne s'en dépossède pas... En découlent les problèmes que pose la prise en compte de l'information en tant que "ressource" dans toute comptabilité à partie double : crédits et débits ne convergent pas et la mise en cohérence des comptes nécessite certains artifices. En outre, celui qui émet une information peut accroître son "stock de connaissance" moyennant un processus de *feed-back* relatif aux modes d'exploitation de celle-ci. Enfin, chaque unité d'information non redondante, contrairement aux objets émanant d'une production en série, est unique : on ne peut échanger la date de naissance d'un citoyen par celle d'un autre. Pour ces raisons, l'évaluation de la valeur monétaire de l'information n'est pas aisée.

Comme le sens de l'information est évolutif en fonction du contexte et des usages, plus celle-ci est partagée par un grand nombre d'utilisateurs avertis la contrôlant fréquemment, dans le cadre d'un suivi organisé, meilleure sera sa qualité. "Use it or lose it !" avance un adage en la matière : les silos d'information peu consultés deviennent souvent obsolètes et incomplets avec le temps (Boydens I., *Informatique, normes et temps*, Bruxelles, Bruylant, 1999). Un des derniers posts (30 juillet 2013) de l'excellent blog en "data quality" de Henrik Liliendahl Sørensen va dans ce sens : "On washing rental cars and shared data".

Cette approche plaide pour les sources authentiques partagées, avec les précautions qu'impliquent les contraintes du domaine

d'application (confidentialité, force probante juridique et contrôle de l'intégrité, par exemple, dans le domaine de l'e-government).

Certains domaines d'application sont en effet particulièrement sensibles...

Si l'on se penche sur le cas de la récente Affaire Edward Snowden, le partage d'information qu'il a orchestré a peut-être permis d'améliorer la qualité des données (même si ce n'était sans doute pas l'objectif essentiel poursuivi), chaque partie concernée dans le monde ayant pu vérifier la validité des renseignements divulgués à propos du Programme Prism et, éventuellement, adapter ses propres données en conséquence ... mais il en a apparemment coûté fort cher à l'intéressé qui, en ce 31 juillet 2013, semble toujours depuis un mois traqué et retranché dans une salle de transit à l'aéroport de Moscou...

- **Il est préférable de livrer en retard un produit qui fonctionne plutôt que de livrer à temps un produit qui ne fonctionne pas... et il en va de même pour l'information**

Cette "bonne pratique" est ancrée depuis longtemps dans la culture industrielle : ainsi peut-on stopper, avant livraison sur le marché, la production de vaccins ou médicaments dont on craint qu'ils soient potentiellement nocifs pour la santé, quitte à faire face à des pertes immédiates, lesquelles sont toutefois moindres que celles liées au risque encouru en cas de mise en production. Il en est de même dans le marché automobile où de grandes marques, telles Toyota, Honda, Nissan ou Fiat ont récemment rapatrié massivement des voitures potentiellement défectueuses (suite à des risques de problèmes d'airbag, de moteurs, de freins, ...). Dans le domaine des systèmes d'information, cette pratique mériterait d'être généralisée, même si la pression des délais et le caractère immatériel des données ne permettent pas de mesurer facilement le terrible impact sur le réel que provoquent inévitablement, un jour ou l'autre, des données inadéquates aux usages (qu'il s'agisse de l'e-government, des matières militaires, scientifiques, environnementales ou médicales).

- **De l'analogie entre processus d'assemblage des produits en usine et processus de production de l'information : le "data tracking"**

Il existe une réelle analogie entre le processus d'assemblage des produits en usine et celui des données au coeur d'un système d'information. Au sein des laboratoires d'AT&T, Thomas Redman a développé sur cette base la technique du "data tracking" en vue de détecter et d'éliminer structurellement à la source les causes de déformation de l'information, en particulier, la redondance, génératrice d'incohérences et d'anomalies fictives. Le "return on investment" d'une telle approche est important. On lutte ainsi contre le syndrome de l'usine fantôme (ou "ghost factory"), désignant le temps et l'argent consacrés par une entreprise à produire des défauts et à les corriger. Dans le domaine de la sécurité sociale, nous avons adapté la technique du "data tracking" et nous l'avons appliquée au cas spécifique de l'e-government : "Améliorer la qualité de l'information : du "stemma codicum" au "data tracking".

- **De par son caractère intangible et immatériel, l'information empirique est évolutive et se transforme avec l'interprétation du réel observable qu'elle permet d'appréhender**

Alors que le caractère défectueux d'un produit industriel se manifeste sous une forme tangible, on ne peut déduire sur la base d'une observation directe si l'information est correcte ou pas. Par exemple, un climat d'humidité peut rouiller une machine métallique et la production de cette dernière peut se réaliser dans des conditions relativement maîtrisées, en vue de prévenir les intrusions néfastes de l'environnement. Par contre, l'information empirique, sujette à interprétation dans le temps évolue dans un environnement ouvert et dynamique. Ce n'est parfois qu'au terme de plusieurs années qu'une inadéquation aux usages apparaît, dans les domaines administratifs, scientifiques ou militaires, lorsque l'information est un instrument d'action sur le réel.

Il en découle que les normes ISO en matière de qualité des données issues des normes industrielles ISO 9000, par exemple, ne sont pas toujours pertinentes. Ainsi, la récente norme internationale ISO 8000 propose, depuis 2009, un Master Data Vocabulary en vue d'améliorer la qualité des données en fournissant des listes de vocabulaire standardisé universel de référence pour assurer une modélisation uniforme de l'information. Le recours à un tel lexique est délicat en raison du caractère des volatile tant du langage naturel que des processus et des réalités

représentés, cela dans tous les domaines empiriques qui sont par essence sujets à interprétations dans le temps et l'espace. Il en est de même de la récente norme ISO 25964 (2011) et (2013) en matière de conception de thésaurus. Celle-ci, en vue de faciliter l'interopérabilité sémantique dans le contexte multilingue du web, repose sur une distinction déterministe entre termes et concepts, comme si le langage était indépendant des réalités qu'il représente ! Une telle approche n'est pas totalement fructueuse sur le plan opérationnel (on lira à cet égard l'excellent mémoire de Laurence Maroye, "*Norme ISO 25964-1 (2011) : apports et limites de la révision des lignes directrices pour la création et la maintenance des thésaurus*", sous la direction d'I. Boydens, Master en Sciences et Technologies de l'Information et de la Communication, Université Libre de Bruxelles, juin 2013). Afin d'évaluer et d'améliorer la qualité des systèmes d'information non structurés et des bases de données, une autre approche reposant sur la prise en compte opérationnelle du processus interprétatif des données est indispensable, ainsi que nous l'avons montré, à titre d'exemple, dans le domaine de l'e-government (voir aussi : Boydens I., *Documentologie*. Bruxelles : Presses de l'Université Libre de Bruxelles, 11ème édition, année 2013-2014/1, 182 p.) .

cost cutting

data quality

egov

information management

information retrieval

managing it costs

MORE POSTS

Je data beschermen tegen beheerders: 'on-premise' Confidential Computing

2026-03

Protéger ses données des administrateurs : l'informatique confidentielle « on-premise »

2026-03

De performance van LLM's: Een vergelijkende analyse tussen Frans en Nederlands

2026-03

Made by Smals Research – Privacyvriendelijk Kruisen van Persoonsgegevens

2026-02

Search

Newsletter & webinars:

 Dutch French

Keywords:

[analytics](#) [artificial intelligence](#) [big data](#) [blockchain](#) [bpm](#) [chatbot](#)[cloud computing](#) [cost cutting](#) [cryptography](#) [data center](#)[data quality](#) [development](#) [eda](#) [egov](#) [event](#) [gis](#)[information management](#) [machine learning](#) [managing it costs](#)[methodology](#) [mobile](#) [natural language processing](#) [open source](#) [privacy](#)[productivity](#) [security](#) [social](#) [software design](#)[software engineering](#) [standards](#)**nals Research**

