

Groupe de contact FNRS

« Analyse critique et amélioration de la qualité de l'information numérique »

28 janvier 2015

# Introduction

Isabelle Boydens



**Smals**  
ICT for society

**fnrs**  
LA LIBERTÉ DE CHERCHER



# Groupe de contact FNRS

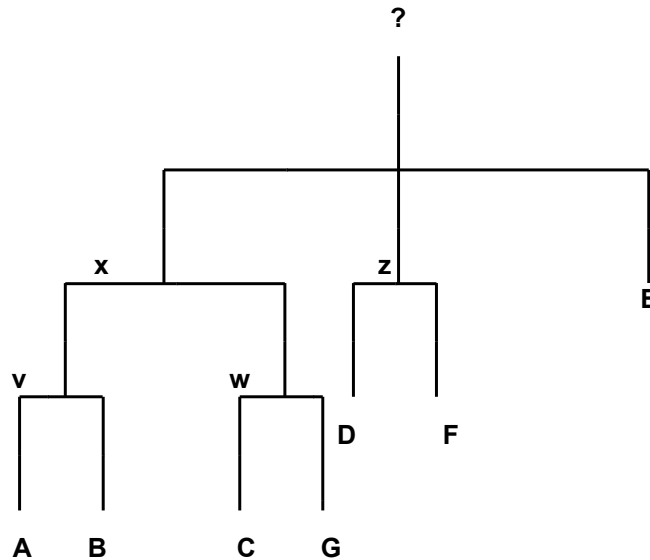
## « Analyse critique et amélioration de la qualité de l'information numérique »

---

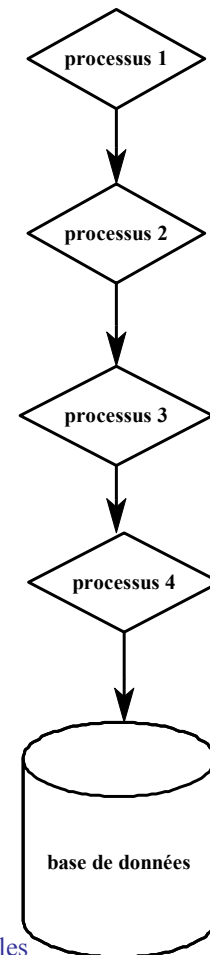
- Qualité de l'information, définition et enjeux (financiers, sociaux, politiques, scientifiques, ...) : « *fitness for use* »
- Groupe FNRS, objectifs, réalisations et défis
  - créé en 1994, pluridisciplinaire
  - associant théorie et pratique, concepteurs et utilisateurs
  - bibliographie en annexe, cours universitaire (stic b 510)
- Application de la critique historique aux sources informatiques à des fins opérationnelles :
  - Herméneutique des bases de données empiriques : « *le système d'information se transforme avec l'interprétation du réel observable qu'il permet d'appréhender...* »

# Du stemma codicum au « data tracking »

**Stemma codicum : "le Lai de l'ombre", poème français  
du 13ème siècle**



**"Data tracking", AT&T Bell  
Laboratories, 1992**



# Enjeux stratégiques lorsque l'information est un instrument d'action sur le réel

---

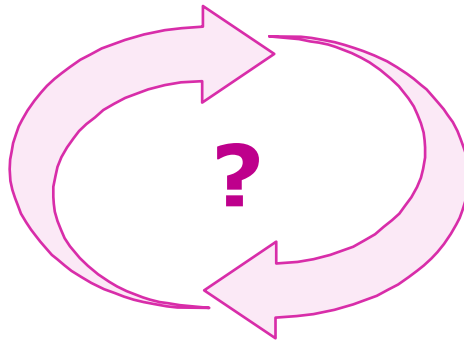
## Quelques exemples :

- 1442, Laurent Valla démontre que la « *Donation de Constantin* » est un faux antidaté de 4 ou 5 siècles
- Étude du réchauffement climatique (couche d'ozone, 1980)
- Première guerre du Golfe (1990-1991)
- 1999, guerre du Kosovo, bombardement de l'ambassade de Chine à Belgrade
- 2008, interactions entre données et marché financier : Google Finance, Yahoo! Finance
- 2013, secteur administratif
- ...

# Les « données » ne sont pas « données »

---

On ne dispose d'aucun référentiel "absolu" en vue de tester  
la correction d'une vaste base de données empiriques



Étude des anomalies à des fins opérationnelles

# Comment les données se construisent-elles progressivement ?

---

## *Cadre d'analyse temporel*

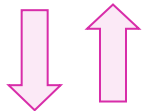
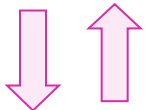
Évolution des normes



Évolution des représentations informatiques



Évolution du réel observable, objet de la norme

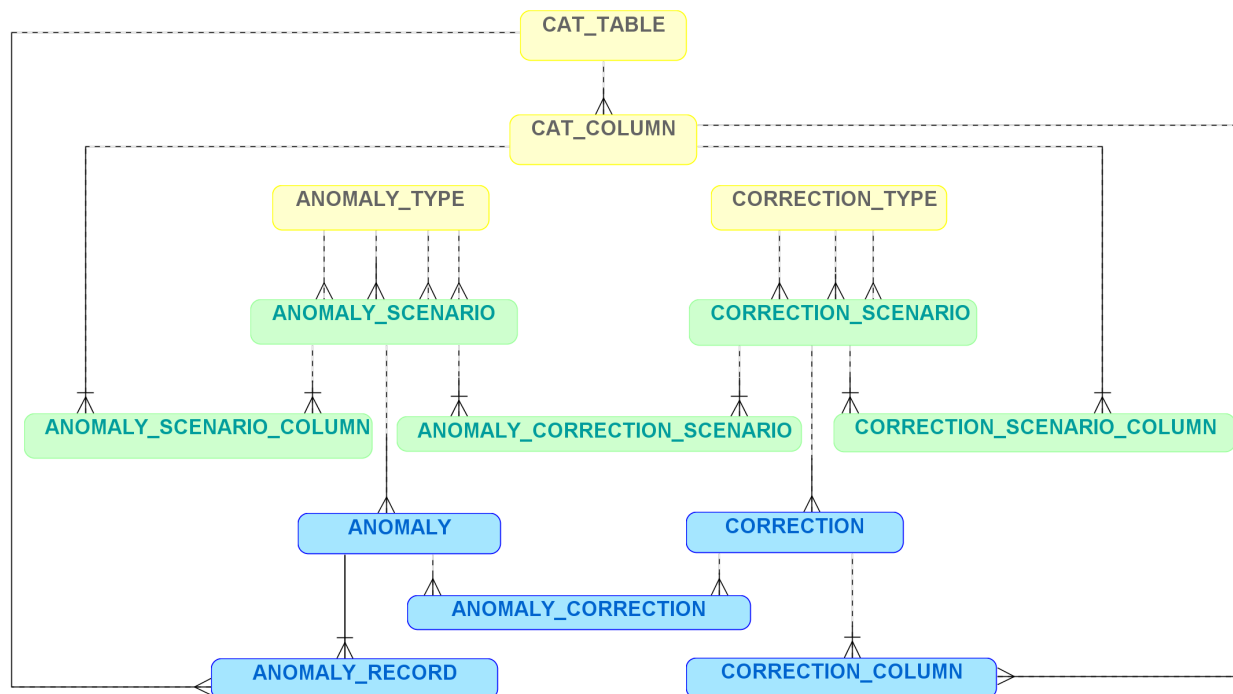


F. Braudel, « *temporalités étagées* » (1976)

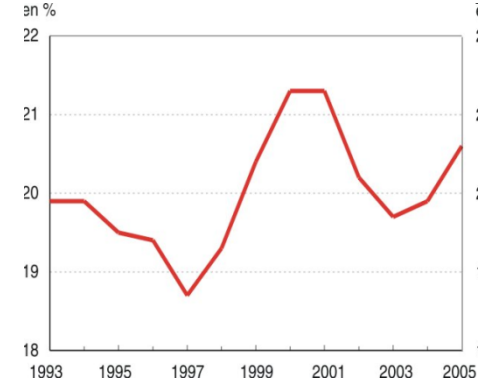
N. Elias, « *continuum évolutif* » (1996)

# Stratégie opérationnelle : passage d'un « monde clos » à un « monde ouvert sous contrôle »

- Extension du modèle de la base données
- Intégration du traitement des anomalies et historique :
  - Typologie et suivi dans le temps
  - Détection / correction / validation ...



# Stratégies de gestion



- Conception d'indicateurs de qualité pour :
  - Quantifier le temps et la nature des traitements (monitoring)
  - Identifier les cas d'anomalies fréquemment validées (anomalies « fictives »)
  - En évaluer les causes (obsolescence des contrôles ?)
  - Adapter progressivement les contrôles pour les rendre adéquats à l'évolution des réalités et diminuer le nombre d'anomalies fictives
  - Exemples d'application concrète (DmfA : 45 milliards d'euros annuels de cotisations sociales prélevées) :
    - Déduction de cotisation pour les "bas salaires"
    - Baisse structurelle du nombre d'anomalies de 50 % (14.000/7.000)
  - Applications dans d'autres domaines empiriques : musées, corpus historiques, secteur bancaire, finances, ...



# Thématique 2015

---

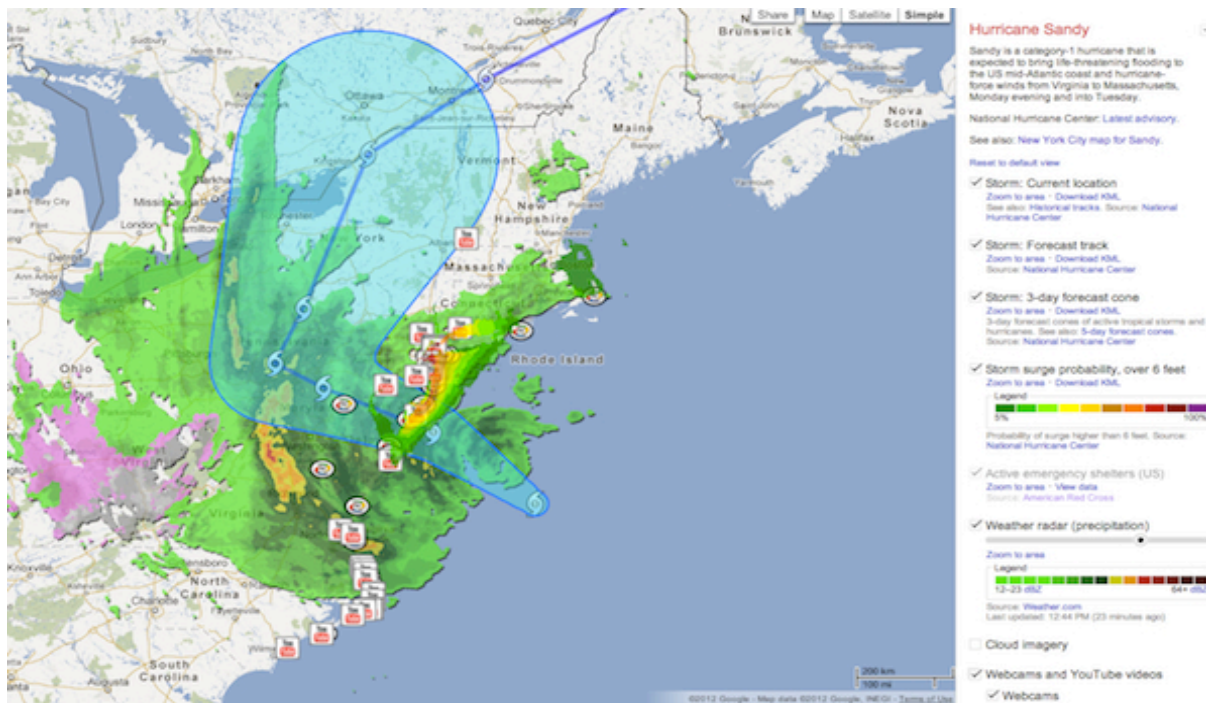
## Gouvernance et engagement citoyen

### mise en perspective philosophique et

examen de la qualité des plateformes européennes de données ouvertes en vue d'une meilleure interopérabilité

- -→ désintermédiation politique: espoirs, limites, perspectives ... Libérer la citoyenneté, comme on croit « libérer » les données ?
- -→ les données ouvertes: une technologie de la démocratie directe ? Qu'en est-il de leur qualité dans un environnement ouvert, hétérogène et multilingue ?

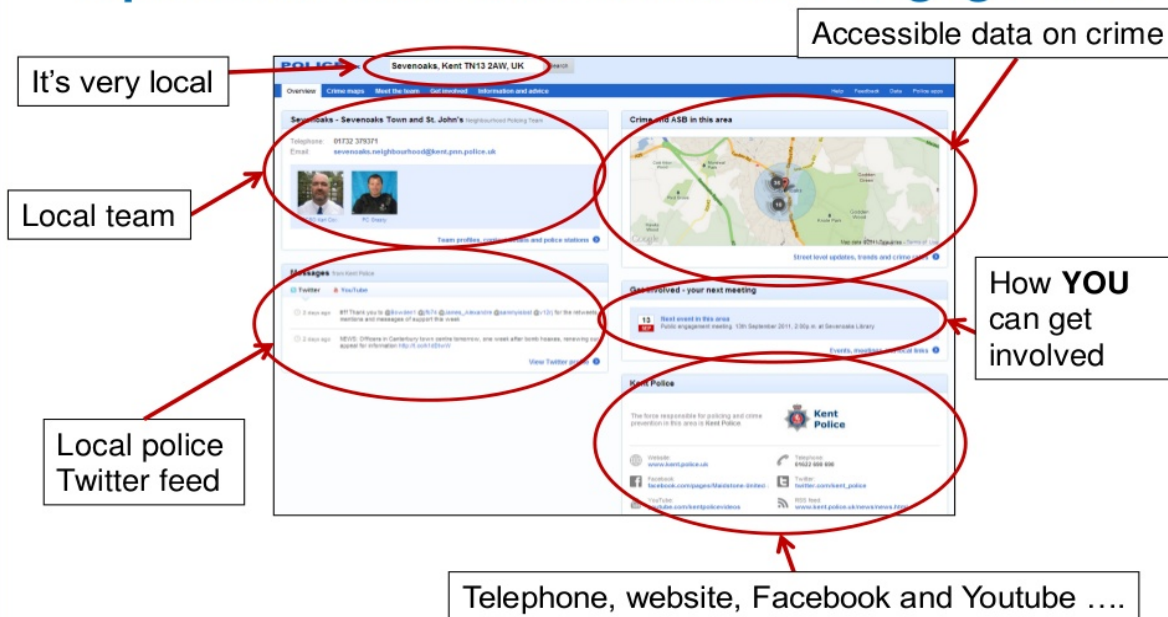
# Suivi real time de l'ouragan *Sandy* (2012)



- données publiques météorologiques, constatations des citoyens, offres de bénévolat, ...
- « citizen engagement », crowdsourcing, ...
- mise en relation de données hétérogènes évoluant à des rythmes distincts

# Citizen engagement and Open Data in UK (Data Days, Ghent, 2014)

## Open Data used to drive Citizen Engagement



19

# Programme 2015

---

- « *Gouvernance et individus: réflexion sur un couple impossible* », par :
  - François De Smet, docteur en philosophie et collaborateur scientifique au CIERL (ULB), <http://francoisdesmet.wordpress.com/>
- « *Semantic Information Technology as a provider of tools for a productive and efficient organization of government ?*», par :
  - Johan De Smedt, Master in Mathematics, Ghent University and CTO TenForce BVBA ([tenforce.com](http://tenforce.com)), involved in many EU projects : **Cellar** (legislation and case law), **ESCO** (the EC labor market taxonomy), the **Open Data Portal**, **Eurovoc** (the EC thesaurus) and the **LOD2** (components for a linked open data platforms).
  - Agis Papantoniou, PhD in knowledge engineering from National Technical University of Athens, Invited Expert at the w3c Government Linked Data Working Group, **ESCO** Program Manager and Senior Project Manager (Semantic Technologies) in Tenforce
  - Bert Van Nuffelen, PhD in Informatics, Department of Computer Science, Faculty of Engineering Science, KU Leuven, **Open DataPortal** and **LOD2** Program Manager, Program Manager in Tenforce
- Table ronde
- Réception

# Orientation bibliographique (1)

---

- Bade D., « It's about Time!: Temporal Aspects of Metadata Management in the Work of Isabelle Boydens ». In *Cataloging & Classification Quarterly* (The International Observer), volume 49, n° 4, 2011, pp. 328-338 (Université de Chicago, recension des recherches et publications d'Isabelle Boydens, période 1993-2011).
- Bade D., *Responsible Librarianship, Library policies for unreliable systems*. Library Juice Press, 2007.
- Berten V. et Boydens I., *Email Address Reliability*, Section Recherches, Bruxelles, Smals, 2014.
- Berti Equille L. éd., *La qualité et la gouvernance des données au service de la performance des entreprises*. Paris, Hermès, 2012.
- Bizingre J., Paumier J. et Rivère P., *Les référentiels du système d'information*. Paris, Dunod, 2013.
- Bloch L. *Système d'information : obstacles et succès*. Paris : Vuibert, 2005.
- Bontemps Y., Boydens I. et Van Dromme D., *Data Quality : tools*. Deliverable, section recherches, Bruxelles, Smals, 2007.

# Orientation bibliographique (2)

---

- Boydens I., « Informatique et qualité de l'information. Application de la critique historique à l'étude des informations issues de bases de données ». In *Belgisch Tijdschrift voor Nieuwste Geschiedenis. Revue belge d'histoire contemporaine*, vol. 3-4, 1993, p. 399-439.
- Boydens I., *Informatique, normes et temps*. Bruxelles : Bruylant, 1999.
- Boydens I., « Les bases de données sont-elles solubles dans le temps? ». In *La Recherche hors série* ("Ordre et désordre"). Hors série n° 9, novembre-décembre 2002, p. 32-34.
- Boydens I., « Déploiement coopératif d'un dictionnaire électronique de données administratives ». In *Revue Document Numérique*, vol. 5, n°3-4, 2001, Paris, Hermès, p. 27-43.
- Boydens I., « La conservation numérique des données de gestion ». In *Revue Document Numérique*, septembre 2004, Paris, Hermès, p. 13-22.
- Boydens I., "Qualité de l'information et administration électronique : enjeux et perspectives". In Assar S. et Boughazala I., eds., *Administration électronique. Constats et perspectives*. Paris : Lavoisier - Hermès Sciences, 2007, p. 103-120 (chapitre 5).
- Boydens I., "Hiérarchie et anarchie : dépasser l'opposition entre organisation centralisée et distribuée ?" In Hudon M. et El Hadi W. M., eds., *Les cahiers du numérique* (Numéro thématique « Organisation des connaissances et Web 2.0 »). Paris : Editions Hermès Sciences, 2010, vol. 6, n°3, p. 77-101.

# Orientation bibliographique (3)

---

- Boydens I., "Strategic Issues Relating to Data Quality for E-government: Learning from an Approach Adopted in Belgium". In Assar S., Boughzala I. et Boydens I., eds., "Practical Studies in E-Government : Best Practices from Around the World", New York, Springer, 2011, p. 113-130 .
- Boydens I. et Van Hooland S., "Hermeneutics applied to the quality of empirical databases". In *Journal of Documentation*, volume 67, issue 2, 2011, p. 279-289.
- Boydens I., Hulstaert A. et Van Dromme D., *Gestion intégrée des anomalies - Evaluer et améliorer la qualité des données*, Delivrible, Section Recherches, Bruxelles, Smals, 2011.
- Boydens I., Mendez E. et Van Hooland S., "Between commodification and sense-making. On the double-sided effect of user-generated metadata within the cultural heritage sector" In Marty P. F. et Kazmer M. M., eds, *Library Trends on "Involving Users in the Co-Construction of Digital Knowledge in Libraries, Archives, and Museums"*, Library Trends, John Hopkins University Press, volume 59, n° 4, spring 2011, pp. 707-720.
- Boydens I., « L'océan des données et le canal des normes ». In Carrieu-Costa M.-J., Bryden A. et Couveinhes P. eds, *Les Annales des Mines, Série "Responsabilité et Environnement"* (numéro thématique : "La normalisation : principes, histoire, évolutions et perspectives"), Paris, n° 67, juillet 2012, pp. 22-29.  
<http://www.ulb.ac.be/cours/iboydens/annales.pdf>
- Boydens I., *Open Data et eGovernment*. Research Note, Bruxelles, Smals, n° 33, avril 2014, 23 pp. [http://www.smalsresearch.be/download/research\\_reports/research\\_note/OpenDataRN.pdf](http://www.smalsresearch.be/download/research_reports/research_note/OpenDataRN.pdf)

# Orientation bibliographique (4)

---

- De Wilde, M. et Verborgh, R., *Using OpenRefine*. Brimingham-Mumbai : Packt Publishing, 2013 (978-1-78328-908-0).
- Elmasri R. et Navathe S. B., *Fundamentals of Database Systems*. Addison Wesley, 2011 (6eme éd.).
- Loshin D., *The Practicioner's Guide to Data Quality Improvment*. Elsevier, Morgan-Kaufmann OMG Press, 2011.
- Madnick S. E. et al., "Overview and Framework for Data and Information Quality Research". In *Journal of Data and Information Quality*, Vol. 1, No. 1, 2009, p. 2-22.
- McCallum Q. Ethan, *Bad Data Handbook, Mapping the World of Data Problems*. Sebastopol, O'Reilly Media, 2012. (analyse critique : <http://blogresearch.smalsrech.be/?p=5398> , I. Boydens, 3 avril 2013)
- Olson J., *Data Quality: The Accuracy Dimension*. Elsevier, The Morgan-Kaufmann Series in Database Management, 2003.
- Redman T. C., *Data Quality for the Information Age*. Boston-London : Artech House Publishers, 1996.
- Redman T. C., *Data Quality. The Field Guide*. Boston : Digital Press, 2001.



# Orientation bibliographique (5)

---

- Rivière P., «Indicateurs de qualité en matière de production de données : quelques éléments de réflexion ». In *Courrier des statistiques*, septembre 2005, n°115, p. 35-40.
- Rivière P., *Les référentiels dans un système d'information. Quelques principes*, Paris, Assurance retraite, Direction Qualité, Méthodes et Urbanisation, 2011.
- Shazia S. ed., *Handbook of Data Quality. Research and Practice*. Berlin, Springer, 2013.
- Van Hooland S., "Spectator becomes annotator: possibilities offered by user-generated metadata for image databases". *Paper presented at Immaculate Catalogues: Taxonomy, Metadata and Resource Discovery in the 21st Century*, 13-15 September 2006, University of East Anglia, UK.
- Van Hooland, S., Kaufman, S. et Bontemps, Y., "Answering the call for more accountability: applying data-profiling to museum metadata", *Proceedings of the International conference on Dublin Core and metadata applications*, 22- 26 September 2008, Berlin, Dublin Core Metadata Initiative, Berlin, p. 93-103.
- Van Hooland, S., *Metadata quality in the cultural heritage sector: stakes, problems and solutions*, Thèse de doctorat sous la direction de Boydens I., Département Sciences de l'Information et de la Communication, Université Libre de Bruxelles., 2009.
- Van Hooland S. et Verborgh, R., *Linked data for libraries, archives and museums. How to clean, link and publish your metadata*. Brimingham-Mumbai : Facet Publishing, 2014.
- Volle M., *De l'informatique*. Paris, Economica, 2006.

# Orientation bibliographique (6)

---

- Quelques ressources en ligne :
  - <http://www.ulb.ac.be/cours/iboydens/>
  - <https://www.smals.be/fr/content/data-quality>
  - <http://www.smalsresearch.be>
  - <http://homepages.ulb.ac.be/~svhoolan/>
  - <http://homepages.ulb.ac.be/~madewild/>
  - <http://freeyourmetadata.org/>
  - <http://liliendahl.com/>
  - <http://iaidq.org/>
  - <http://www.ocdqblog.com/>
  - <http://www.dqa.be/>
  - <http://exqi.asso.fr/>

