



# Hermeneutics applied to the quality of empirical databases

Hermeneutics  
applied to  
quality

Isabelle Boydens

*Information and Communication Science Department,  
Université Libre de Bruxelles, Brussels, Belgium, and*

Seth van Hooland

*Library and Information Science Department,  
Universidad Carlos III de Madrid, Madrid, Spain*

279

Received 12 December 2009  
Revised 16 March 2010  
Accepted 29 May 2010

## Abstract

**Purpose** – This paper seeks to present a conceptual framework to analyze and improve the quality of empirical databases throughout time – with operational results which are measurable in terms of cost-benefit.

**Design/methodology/approach** – Basing themselves on the general approach of hermeneutics and, more specifically, on Fernand Braudel's concept of "*temporalités étagées*" and Norbert Elias's "evolutive continuum", the authors develop a temporal framework consisting of three stratified time levels in order to interpret shifts in the quality of databases. The soundness of the framework and its capability of delivering operational results are demonstrated by the development of a case study focusing on social security databases. A second case study in the context of digital cultural heritage is also developed to illustrate the general applicability of this interdisciplinary approach in the context of empirical information systems.

**Findings** – Contrary to the assertions of common theories that postulate a permanent bijective relationship between records in a database and the corresponding reality, this paper provides insights which demonstrate that a database evolves over time along with the interpretation of the values that it allows one to determine. These interdisciplinary insights, when applied practically to concrete case studies, give rise to original operational results in the ICT field of data quality.

**Practical implications** – The framework helps both the managers and the users of empirical databases to understand the necessity to integrate unforeseen observations, neglected *a priori* by virtue of the closed world assumption, and to develop operational recommendations to enhance the quality of databases.

**Originality/value** – This paper is the first to show the potential of hermeneutics for the task of understanding the evolution of an empirical information system, and also the first to deliver operational outcomes.

**Keywords** Databases, Information management, Quality, Interpretive programs

**Paper type** Conceptual paper

## 1. Introduction

The impetus for this paper is the pertinence of heuristics, the "historical method" and hermeneutics in the context of Computer and Information Science applied to the management of empirical databases. These traditional methodologies from the humanities, which have been used as scientific tools for over a century, can still be effectively applied even if these methods need to be adapted to the increasingly complex and interdisciplinary nature of contemporary application domains (see, for example, Aron, 1969; Dilthey, 1988).



The sense-making process of interpreting individual empirical events comes about by placing the singular events in a more abstract and generalizing framework. Heuristics and hermeneutics offer the tools to create such a unifying and sense-making horizon. This framework is typically created through a continuous process of going back and forth between the individual observations and the larger context in which the observations can be framed. We can for example refer to the medical domain, in which theories evolve along with live experience, as demonstrated by the current research regarding the A H1N1 flu. This continuous process of matching the evolution of reality with updates of the sense-making framework also lies at the heart of our research on the quality of empirical databases.

Hermeneutics, and particularly the temporal framework as developed by Fernand Braudel, have been applied throughout various application domains. Only a limited number of authors from the Information Science domain, such as Boyd Rayward (1996) and Ribes and Finholt (2009), have applied the temporal framework to their research. However, both authors limit the use of the framework to a level of analysis, and do not demonstrate how operational results can be obtained, nor do they enrich Braudel's framework with Norbert Elias's theory, for which time is an arbitrary construction resulting from the relationship between two or more sequences of transformation. It follows that their approach is not truly interdisciplinary, nor applied to concrete case studies.

With the help of two largely differing case studies, this paper demonstrates how hermeneutics, applied to the temporal framework, can lead to operational results regarding the quality of empirical databases. After an outline of the strategic importance of data quality, section 2 gives an overview of the existing approaches which have been developed to tackle this issue and finally defines the originality of the hermeneutic approach we propose for data quality.

The next section, which constitutes the core of the paper, develops in detail the concept of the temporal framework, based on the work of Fernand Braudel and Norbert Elias. A case study of Belgian social security databases, which manage the collection and redistribution of some 40 billion euro every year, is used to elaborate and illustrate the theoretical framework. These administrative databases store quarterly approximately 4,000,000 records, with several hundred corresponding attributes. Every quarter, hundreds of thousands of formal anomalies are detected, due to the ever-evolving nature of the reality that the database records aim to represent. Therefore, this application domain offers a suitable and strategic case to illustrate the use of the temporal framework coming to grips with data quality in empirical databases.

The second half of the third section introduces a second case study from a different application domain, the cultural heritage sector, in order to exemplify the general approach of the temporal framework in a wide variety of application domains. Because of the very recent and still evolving nature of its holdings, the collection management database of the September 11th Memorial and Museum in New York struggles to find a stable and appropriate metadata scheme to adequately describe its collection. The case study reports on the development of a prototype which has been implemented into the collection management database to monitor the use of the different metadata fields over time, and thus to detect evolutions in the use of the different available metadata fields.

The conclusions offer a critical evaluation of the temporal framework and point out to future research perspectives.

---

## 2. Data quality and hermeneutics

The quality of databases, and of the data they contain, is considered today a strategic matter. The question is of primary importance when the data are used as a tool to assist decision making, or even real-world action. For instance, in May 1999, during the war in Kosovo, NATO mistakenly bombed the Chinese embassy in Belgrade: the mapping databases that were used to guide the missiles contained a plan of the city which was obsolete and therefore inadequate. Hence, the untimely attack and the diplomatic incident which followed (Boydens, 2007).

In database modeling theory (Elmasri and Navathe, 2007), a database is composed, on the one hand, of a structure (or “database schema”) that specifies, on the basis of a set of integrity constraints, its semantically allowed content and, on the other hand, of a set of records (“database extension”) or of data stored in the database and satisfying its structure at a given moment. The critical focus of this paper considers the interactions between the semantics of the database schema and the database extension. But how can one define data quality? We can start off with the ISO 9000 definition of quality, which describes it as the “totality of features and characteristics of a product, process or service that bears on its ability to satisfy stated or implicit needs” (ISO, 2005). Therefore, the quality of a database denotes its adequacy with respect to the purposes assigned to it, which can be referred to as the “fitness for use” principle. “Total quality” does not exist, since the concept is relative: on the basis of a cost-benefit type analysis, the most pertinent quality criteria – that can be the timeliness of information, the speed of data transmission or of user access — must be adopted in a given context.

Starting from the 1990s, different strategies have been developed to approach the problem of data quality in information systems. Thomas Redman has developed a system-centered method in which three complementary analyses build a framework for the management of database quality: data modeling, data values and data representation (Redman, 1996). A mixed approach between the systems- and a more user-centered strategy was initiated in the beginning of the 1990s at the MIT, with the research of Mark Hanzen and the development of the total data quality management (TDQM) program of Richard Wang (Wang and Wang, 1996).

Distancing herself quite radically from these two research centers, Boydens (1999) has developed a hermeneutic approach toward database quality. Although Redman suggests an interesting model that can be applied to a wide variety of contexts, he only deals with errors than can be formally identified and measured as such, such as discrepancies and incompleteness (see for instance, the “data tracking” model). But nowhere does Redman deal with the difficulty of human interpretation of information. Taking her critique a step further, Boydens (1999, p. 9) especially denounces the data quality research from the TDQM program, it being based on the assumption that it is possible to detect a formal error within a database by comparing the content of the database with the reality the database strives to represent:

How can we know whether data are valid? The answer seems overly simple: in order to evaluate the correctness of values within a database, we should compare them with the reality the values represent. However, this line of thinking is underpinned by the implicit hypothesis that there is a permanent bijective relationship between data and the corresponding reality. We argue that this type of isomorphic relation does not exist in empirical application domains.

Empirical information systems evolve over time, along with the interpretation of the values that they allow to determine. Based on the hypotheses of the TDQM program, information quality formulas have been developed, which propose to calculate, for instance, the required level of correction of a database regarding the reality presented in it (Madnick *et al.*, 2009). Obviously, these formulas have never been put into practice within the industry.

Currently, a set of data quality tools, based on data profiling techniques – standardization, matching and cleansing – are being effectively applied within the corporate world. These are all useful techniques, but they do not include the interpretative aspect of our approach, which is vital in the context of empirical databases. The technique of data cleansing, for example, involves the automated “smoothing” of the content of a database *a posteriori* by eliminating the values considered to be aberrant by using a set of formal correction algorithms. While this technique is valuable in the statistical domain, where a certain error rate is permissible, it is not valid in the administrative domain, where each individual case has to be considered (Olson, 2003). We also insist that data cleansing does not act on the causes of the “no-quality”.

To tackle the issue of data quality at a more fundamental level, one needs to clearly distinguish deterministic data from empirical data. The first are characterized by the existence of a persistent theory which makes it possible to decide whether or not a value is correct. This is the case with algebraic data: since the rules of algebra do not change over time, we can know at any time whether or not the result of a sum is correct. For empirical data, which are subject to human interpretation, theories change over time along with the interpretation of the values that it has made it possible to determine. This is true in the economic domain (e.g. with regard to the calculation of national wealth), but also in the legal and administrative fields, where the interpretation of legal concepts steadily changes with the constant evolution of real-world circumstances and jurisprudence. For example, when “copy centers” (i.e. shops offering photocopying services to their customers) first began to appear, the nomenclature for European business activities (used in administrative databases in order to categorize companies) was quickly found to be inadequate for their classification: the best it could offer were the categories of “printing”, “book retailing” or “secretarial services”. To take the category “copy centers” into account, it was first necessary to amend the regulatory texts, and then to adapt the structure of the administrative databases accordingly. This remains crucial in the context of electronic administration. For example, the notion of “principal activity” of a company, which is fundamental to business registers, is a mutable factor whose reliability is difficult to evaluate. These problems are also encountered on a daily basis by labor inspectors doing fieldwork:

In almost 30 years of doing this job [...], I have seen many points become subject to interpretation, and therefore to challenge. It's a little like a policeman stopping you for going too fast and telling you that you're driving at a “dangerous” speed. You could easily contest this notion (Baroux, 2006, p. 3).

Thus, to address empirical data quality (i.e. “fitness for use”), the structure of the database has to evolve continuously along with the changes in the corresponding observable reality. We will now consider how to apply this mechanism using our temporal framework.

---

### 3. “*Temporalités étagées*“ and the mapping of database records with the evolving reality

Having demonstrated the impact of the evolution throughout time of the real world on the quality of a database and its records, which try to represent the same reality, we present a framework that helps to detect these evolutions and to develop operational recommendations.

#### 3.1 *Understanding the temporal framework*

Different approaches have been developed from a hermeneutic point of view in order to analyze and come to terms with heterogeneous evolutions that influence the same reality. In his book *The Mediterranean and the Mediterranean World in the Age of Philip II*, the French historian Fernand Braudel developed the concept of the temporal time, in order to establish a hierarchy between different co-existing transformations of reality. He distinguishes the long term of geographical structures, the middle term of economical conjectures and the short term of political events. As Antoine Prost explains, this abstract framework allows us to:

[...] distinguish different but interacting time levels, just like the depth of field of a lens allows a film maker to show different characters within the same scene, but all individually distinct, as they are at different distances from the camera lens (Prost, 1996, p. 118).

The originality of our approach resides in an enrichment of Braudel’s original framework, allowing the use of the temporal framework within the context of database quality to obtain operational results. Within Braudel’s framework, the slower sequences have an impact upon the faster ones, in the sense that geographical characteristics influence the economical development, for example. However, this static and uni-directional influence does not match the empirical observations described in the two case studies presented in sections 3.2 and 3.3, as we can also detect influences in the opposite direction. Referring to the example mentioned above, the arrival of copy centers, which is an evolution of the observable reality we place within the short term (see below), has an impact on the administrative and technological apparatus (medium term) and on the legislation (long term) used to describe and manage the observable reality.

In order to acknowledge this phenomenon, we use Elias’ concept of the “evolving continuum”, with which Elias demonstrated that time is a construction resulting from the relationship between two or more sequences of transformation. Each development belongs to the “evolving continuum”, in which one sequence of transformation is normalised in order to serve as a bench mark for the other transformations (Elias, 1996). The following two case studies will demonstrate how this enriched framework allows us to obtain operational outcomes.

#### 3.2 *Applying the temporal framework to the context of social security databases*

Braudel’s temporal framework, enriched with Elias’s evolutive continuum, can be applied to the field of administrative databases. The three levels of transformation originally developed by Braudel are interacting in the information system: the evolution of jurisprudence, the changes made within databases and the categories observable in the field. These three levels of reality are interlinked, but asynchronous as they operate, according to their nature, on different timescales. Thus we have the “long term” for legal rules, renewed from one quarter or one year to the next, the

“medium term” for the management of databases, and the “short term” for the observable reality, i.e. that of the citizens or companies subject to administration, which is continuously evolving. Companies regularly merge, split or disappear altogether, while new professions or categories of activity not covered by the official nomenclatures constantly come into existence, as with the diversification of IT jobs.

From a dynamic point of view, an ideal database should therefore match the rhythm of its updates to the (unforeseeable) division into “layered timescales” of the changes in reality that it seeks to grasp. We must add to this the necessity, always revealed a posteriori, to integrate unforeseen observations, prohibited a priori by the closed world assumption, which states that, for evident operational reasons, every fact that violates the database formal integrity constraints is considered to be false (Elmasri and Navathe, 2007; Boydens, 1999).

Let us take an example from the domain of employment creation measures. Following the directives issued by the European Council in Brussels in December 1993, on the basis of Jacques Delors’ White Paper on growth, competitiveness and employment, job creation measures have multiplied in most European countries in order to fight unemployment. Thus in Belgium, during the implementation of an administrative guideline aimed at the “non-market” sector, the question arose, with regard to the reality progressively reflected in the database, of whether this sector should include private nursing homes, which were a priori excluded because they operate for profit. Initially regarded as “erroneous” cases with respect to the domain of definition for the “non-market” sector, these businesses were eventually included after legal interpretation. This led to a restructuring of the database schema, which was the result of a human decision aimed at bringing the model temporarily in line with the new observations. The non-market category has actually been restructured both in the database schema and in the law in order to include the “private nursing homes”. This operation was applied with a retroactive effect of more than one year (15 months: the law modified on July 1, 1998 had to be applied from April 1, 1997) and the previously stored database extensions were enriched with the companies belonging to the “private nursing homes” category, previously considered as “anomalies”. This phenomenon of transformation corresponds to the so-called “strange loop” mechanism defined by Hofstadter (1980). In the absence of such an intervention, the gap between the database and reality widens.

What are the consequences of this analysis with regard to specifying appropriate quality indicators for administrative information? Since administrative data are empirical by nature, there is no direct frame of reference for testing its correction. Their appropriateness to the needs of the field can be determined only indirectly, via a series of lateral indicators. First, it is necessary to consider the relative pertinence of the information with respect to the objectives pursued: pertinence is a non-quantifiable indicator. Next, a series of quantifiable indicators related to the detected anomalies and their handling may be produced with a view to deploying management strategies for the database.

Finally, in all cases, there must be a tool for the critical interpretation of the data. On the basis of Braudel’s and Elias’s analyses, applied to administrative databases and of a hermeneutic approach, we present an example of the operational exploitation of the obtained indicators. Statistical monitoring of integrity constraints violations (“formal anomalies”) makes it possible to detect not only “abnormal” increases in anomalies

---

(with respect to a given threshold), but also increases in “validations” of anomalies during the handling phase. A validation operation means that, after examination, an operator has judged that the anomaly, which is a presumption of error, corresponds to a pertinent value. The operator can coerce the system into accepting the value. If the rate of these anomaly validations is high and recurring, there is a high probability that the structure of the database itself is no longer pertinent. An algorithm then issues a “signal” to the database manager so that he or she can consider whether a structural modification of its schema is required. When there are large numbers of validations, it is worthwhile to scrutinize the phenomenon closely: as we have seen, a new circumstance (e.g. the emergence of a new category of activity or a change in the interpretation of a concept, such as the “non-market” sector cited above) may have arisen, requiring a modification of the database structure. If the schema is not modified accordingly, the anomalies corresponding to these cases will continue to appear in large numbers, demanding a potentially large-scale manual examination and considerably slowing down the administrative dossier handling. For the Belgian social security system, the implementation of this method made it possible to improve the precision and speed of social security contributions handling, reducing by up to 50 p.c. the volume of formal anomalies, which had previously accounted for 100,000 to 300,000 occurrences to be managed manually every quarter (Boydens, 1999). Other types of indicators for identifying, quantifying and categorizing anomalies and the nature of their handling are essential for the implementation of efficient electronic administration services.

### *3.3 Applying the temporal framework to the context of cultural heritage databases*

Now that the application of the temporal framework has been demonstrated in the context of the management of social security databases, a second and widely differing case study is proposed to illustrate the possibility of using the same temporal framework in an entirely different application domain. The cultural heritage sector has embraced the use of databases to manage digitized collections and the metadata that describe the collection items. The economical and social stakes within the cultural sector are of a different order compared to the social security sector, but cultural heritage institutions are currently also facing data quality issues (Van Hooland, 2009).

The temporal framework described above can be effectively used to distinguish the different evolutions that have an impact on the quality of metadata regarding cultural heritage resources. Within the “long term”, we can place the evolution over decades of the government policies toward cultural heritage preservation and access, including the shifts in the perception of what should be considered cultural heritage and how the public can or should relate to that heritage. The promotion of metadata creation by laymen is a good illustration of how policy changes can affect metadata quality (van Hooland, 2006). The “medium term” gathers the evolution of technologies used for metadata creation, evolving over a three to five year basis. The shift from stand-alone to web-based collection management software has for example, created more possibilities for distributing the metadata creation process. Finally, the “short term” analyses the evolution of the metadata themselves, which are the object of the policies and the technological apparatus.

Cultural heritage data are, as it was the case with the administrative data from the previous case study, empirical by nature, and equally lack a direct frame of reference

for testing their correction. Their appropriateness to the needs of the field can be determined only indirectly, by considering the relative pertinence of the information with respect to the objectives pursued. Faced again with the evolutions that operate on different timescales, the need rises to match the shifts in the use of cultural heritage resources with updates of the collection management database and the metadata scheme. In order to acknowledge the real-life impact of this theoretical observation we have developed and implemented a prototype of a dynamic interface allowing the monitoring of the day-to-day use of the metadata within a collection management database (Van Hooland, 2009; Méndez and van Hooland, 2009). Concretely, the idea behind the dynamic interface is that the user is confronted with a default search interface which can be configured intuitively for the individual needs of the user. Inspired by the way blocks of information can be “dragged and dropped” on the iGoogle interface, the application gives the possibility to customize the search interface by adding, deleting and rearranging the different metadata fields. Behind the direct interest for the database users of having a customizable search interface lies the possibility for the collections manager to monitor which metadata fields are actively being used within the database. The use data are collected in a seamless and automated way and allow the monitoring of changes in user needs over time.

The prototype of the dynamic interface has been developed in close collaboration with the development team of CollectiveAccess, a highly configurable open source collection management system. The National September 11th Memorial and Museum in New York, which uses this software package to manage objects and digital assets related to the events of September 11, 2001, was asked to act as beta users of the dynamic interface. The highly evolving and very recent character of the National September 11th Memorial and Museum makes its collection a perfect case study to analyze the adequacy of the dynamic interface tool. New object types enter the collection frequently and require the collection managers to rethink the existing metadata scheme. When new object types such as destroyed fire trucks or steel construction pillars from the World Trade Center enter the collection, the existing metadata scheme is not necessarily capable of describing these new types of collection holdings. Not only are there still new types of objects emerging from the World Trade Center site itself, but the knowledge about 9/11 and its impact continues to develop under the influence of current and future political events. New forms of remembrance and dealing with this event will therefore need to be incorporated within the collection management database. The implementation of the prototype of the dynamic search interface can monitor how the search behavior evolves in this context by analyzing which specific metadata fields are actually used and which are not.

Collection managers will be able to use the outcomes from the dynamic interface to judge the relevance of the different metadata fields, and provide statistical backing when deciding on resource distribution for the creation and maintenance of metadata. As the case of use evaluation of electronic resources within the academic library world has demonstrated, the linking of statistics with management decisions is not straightforward and needs to be carefully put within a larger perspective (Best, 2007). Nevertheless, metadata providers need to start experimenting with tools and methodologies which allow them to monitor the effective use of the metadata they produce by its user community. The dynamic interface presented above offers a first essential step in this direction.

---

Based on the analysis of the log files, collection managers will also have the opportunity to assess the adequacy of the metadata scheme they are currently using to describe and access their holdings. Statistical analyses of the search fields which are added or deleted from customized search forms will point out the relevance of the individual metadata fields. Collection managers within the cultural heritage sector are very often confronted with backlogs of holdings that have not been described due to a lack of resources. The results from the dynamic search interface may give these managers an automated tool to guide their decisions regarding the optimization of the metadata scheme. Basing themselves upon statistical analyses, collection managers can for example decide to prioritize work on the most used metadata fields and have the initial metadata schema being adapted accordingly.

The dynamic interface is still in a prototype phase, but this tool will be included as a permanent feature within the next version of the CollectiveAccess software, which is used by a wide variety of heritage institutions throughout Europe and the USA. The distribution of the dynamic interface tool throughout different types of heritage institutions will allow us to aggregate towards the summer of 2010 use data from a wide range of metadata providers.

#### 4. Conclusion

This article has illustrated how hermeneutics, embodied through the use of a temporal framework, can help to interpret changes in the quality of empirical databases and lead the way to operational recommendations. By doing so, we have adopted a radical interdisciplinary approach through the confrontation of interpretative theories from the humanities with operational theories in the field of database modeling. This confrontation contributes to the analysis of the iterative interaction between the schema of an empirical database and its corresponding extension and to the formulation of original operational solutions to improve database quality, i.e. its adequacy to observable reality and uses.

In conclusion, we would note that phenomena similar to those observed in the administrative and cultural sector are found in every empirical domain. This is true, for example, in the world of stratospheric databases: before the discovery of falling ozone levels by British researchers in the 1980s, the corresponding low values were, as a matter of course, treated as anomalies (errors due to mis-measurement) in NASA's database for more than a decade (Boydens, 1999). The prevailing theory of the time, which was modeled in the NASA database, did not allow to entertain the possibility that such values might be correct. After the British discovery, NASA adapted the structure of its database, integrating the values previously regarded as errors due to mis-measurement into the set of permissible values. The management strategy described in this chapter therefore applies to all information systems whose structure evolves according to the interpretation of the realities that they aim to grasp. This is particularly true of empirical databases, in which the homogeneity of the formal codifications clashes with the heterogeneity of the empirical categories.

Such a mechanism has been taken into account in the field of thesauri, as the ISO Standard (see for instance: ISO (1986)) includes the notion of "candidate descriptor", which takes into account potential new concepts corresponding to an evolution of the observed empirical reality. Nevertheless, in the case of thesauri, the discovery of new categories requires most of the time a pure intellectual analysis, even if semi-automatic

approaches are under investigation (see, for instance, Schneider and Borlund (2005)). In the field of large databases, we show that a semi-automatic mechanism can be applied to this end on the basis of the monitoring of anomalies treatment or log files about metadata categories. Moreover, this phenomenon is not yet taken into account in the field of database modeling and data quality.

In the future, we aim to pursue our work in the field of data quality by addressing some current shortcomings of our research and to deepen our understanding of the problem area. We have applied our analysis with the temporal framework in the field of administrative and cultural databases. In order to strengthen our research outcomes, both on a theoretical and an operational level, other application domains such as the medical, archaeological, military, etc., are envisioned.

The paper mentioned the usefulness of existing data quality tools and methodologies to detect and correct anomalies in databases, which can then lead to updates of the database structure. We aim to work on a set of best practices, describing an optimal combination of an automated and a manual-interpretative approach to tackle data quality within empirical databases. The blend of automated and interpretative methods is particularly essential when confronted with large data sets.

The long-term implementation of the strategies we propose requires the creation of a knowledge base containing documentation on the interpretation and the correction of values, but also of the restructuring process of the database and/or the metadata scheme. The implementation process of such a knowledge base needs to be formalized in order to automate certain parts of its construction and maintenance.

## References

- Aron, R. (1969), *La philosophie critique de l'histoire*, Vrin, Paris.
- Baroux, R. (2006), "En première ligne sur le front du travail", *Le Monde*, 16 March, p. 3.
- Best, R. (2007), "Lies, damn lies and usage statistics: what's a librarian to do?", in Fowler, D. (Ed.), *Usage Statistics of E-serials*, Haworth Information, Binghamton, NY, pp. 199-214.
- Boyd Rayward, W. (1996), "The history and historiography of information science: some reflections", *Information Processing and Management*, Vol. 32 No. 1, pp. 3-17.
- Boydens, I. (1999), *Informatique, normes et temps*, Bruylant, Brussels.
- Boydens, I. (2007), "Qualité de l'information et e-administration: enjeux et perspectives", in Assar, S. and Boughazala, I. (Eds), *Administration électronique: constats et perspectives*, Lavoisier-Hermès Sciences, Paris, pp. 103-20.
- Dilthey, W. (1988), *L'édification du monde historique dans les sciences de l'esprit*, Editions du Cerf, Paris.
- Elias, N. (1996), *Du temps*, Fayard, Paris.
- Elmasri, R. and Navathe, S. (2007), *Fundamentals of Database Systems*, Addison-Wesley, Reading, MA.
- Hofstadter, R.D. (1980), *Gödel, Escher, Bach, an Eternal Golden Braid. A Metaphorical Fugue on Minds and Machines in the Spirit of Lewis Carroll*, Penguin Books, New York, NY.
- International Organization for Standards (ISO) (1986), *ISO 2788:1986 Guidelines for the Establishment and Development of Monolingual Thesauri*, available at: [www.iso.org/iso/catalogue\\_detail.htm?csnumber=7776](http://www.iso.org/iso/catalogue_detail.htm?csnumber=7776)
- International Organization for Standards (ISO) (2005), *ISO 9000:2005: Quality Management Systems – Fundamentals and Vocabulary*, ISO, Geneva.

- 
- Madnick, S.E., Wang, R.Y., Lee, Y.W. and Zhu, H. (2009), "Overview and framework for data and information quality research", *Journal of Data and Information Quality*, Vol. 1 No. 1, pp. 2-22.
- Méndez, E. and van Hooland, S. (2009), "Metadata typology and metadata uses", in Sicilia, M.A. (Ed.), *Handbook of Metadata, Semantics and Ontologies*, World Scientific Publishing, London.
- Olson, J. (2003), *Data Quality: The Accuracy Dimension*, Morgan Kaufmann, San Francisco, CA.
- Prost, A. (1996), *Douze leçons sur l'histoire*, Seuil, Paris.
- Redman, T. (1996), *Data Quality for the Information Age*, Artech House, Boston, MA.
- Ribes, D. and Finholt, T. (2009), "The long now of technology infrastructure: articulating tension in development", *Journal of the Association for Information Systems*, Vol. 10, special issue, pp. 375-98.
- Schneider, J. and Borlund, P. (2005), "A bibliometric-based semi-automatic approach to identification of candidate thesaurus terms: parsing and filtering of noun phrases from citation contexts", *Proceedings of the 5th International Conference on Conceptions of Library and Information Sciences, CoLIS 2005, Glasgow, UK, June 4-8, 2005*, Springer, Heidelberg, pp. 226-37.
- Van Hooland, S. (2006), "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, University of East Anglia, 13-15 September 2006, available at: <http://homepages.ulb.ac.be/~svhooland/Usergeneratedmetadata.pdf> (accessed November 25, 2009).
- Van Hooland, S. (2009), "Metadata quality in the cultural heritage sector: stakes, problems and solutions", PhD thesis, Université Libre de Bruxelles, Brussels.
- Wand, Y. and Wang, R.Y. (1996), "Anchoring data quality dimensions in ontological foundations", *Communications of the ACM*, Vol. 39 No. 11, pp. 86-95.

### Further reading

- Boydens, I. (2010), "Strategic issues relating to data quality for e-government: learning from an approach adopted in Belgium", in Assar, S., Boughzala, I. and Boydens, I. (Eds), *Practical Studies in E-government*, ISTE, Eugene, OR.
- Braudel, F. (1976), *La Méditerranée et le monde méditerranéen à l'époque de Philippe II*, Armand Colin, Paris.
- Redman, T. (2001), *Data Quality: The Field Guide*, Digital Press, Boston, MA.
- Van Hooland, S., Kaufman, S. and Bontemps, Y. (2008), "Answering the call for more accountability: applying data-profiling to museum metadata", *Proceedings of the International Conference on Dublin Core and Metadata Applications, Berlin, 22-26 September 2008, Berlin*, Dublin Core Metadata Initiative, pp. 93-103.

### Corresponding author

Isabelle Boydens can be contacted at: [iboydens@ulb.ac.be](mailto:iboydens@ulb.ac.be)