Answering the Call for more Accountability: Applying Data Profiling to Museum Metadata

Seth van Hooland
ULB, Belgium
svhoolan@ulb.ac.be

Yves Bontemps
IBM, Belgium
yves.bontemps@be.ibm.com

Seth Kaufman
OpenCollection, USA
seth@opencollection.org

Abstract

Although the issue of metadata quality is recognized as an important topic within the metadata research community, the cultural heritage sector has been slow to develop methodologies, guidelines and tools for addressing this topic in practice. This paper concentrates on metadata quality specifically within the museum sector and describes the potential of data-profiling techniques for metadata quality evaluation. A case study illustrates the application of a general-purpose data-profiling tool on a large collection of metadata records from an ethnographic collection. After an analysis of the results of the case-study the paper reviews further steps in our research and presents the implementation of a metadata quality tool within an open-source collection management software.

Keywords: metadata quality; data-profiling; collection management software

1. Introduction

Collection registration technologies for cultural heritage resources have greatly improved during the last three decades, gradually transforming card catalogs to web-based applications. Successive technologies have impacted the content of both newly created metadata and existing metadata migrated from older platforms. A good example of the influence of a specific technology on content is the character field length limitations of punch cards fed into mainframes in the 1970's, the effects of which are still felt today in some legacy data sets. Technological evolutions have also been accompanied by (and partially engendered) a shift in the profile of professionals working with these tools to document collections. There is, for example, a clear tendency within cultural institutions to give the repetitive work of metadata creation to administrative and technical staff, apprentices or student workers, whereas collection description used to be performed by specifically trained staff members. In multi-lingual countries such as Belgium one also has to consider the complexity of collections being described sometimes in one language, sometimes in another, depending on the mother tongue of the staff. Under these circumstances vast repositories of metadata records have been created and migrated from one platform to another, with little or no information regarding their consistency, completeness and accuracy.

As long as the metadata remained within the safe boundaries of the museum this was not such a problem. Users submitted their question to a member of the museum staff that could query the database for them. As such, the database (and the metadata records it contained) was more or less treated as an internal tool. But then came the web. Initially, most museum web-presences were limited to basic institutional information. Only a very limited number of museums published their metadata in the same way as libraries, which offered their users an OPAC. But the growing tendency to aggregate thematically or geographically related metadata from libraries, archives and museums with the use of OAI-PMH has raised the pressure on museums to publish or distribute all of their available metadata. The disappointing quality of search results and the minimal descriptions attached to retrieved objects within such projects has led to a discussion on issues surrounding the consistency, accuracy and completeness of metadata.

This discussion is badly needed as collection holders increasingly try to re-use metadata and gain more value from them within digitization projects. Metadata practitioners assisting

digitization projects that aggregate metadata of different partners must acknowledge that the quality of existing metadata is hardly questioned. After all, which collection holder wants to stand up in the middle of his or her peers and warn them about the poor quality of his or her metadata?, This misplaced trust causes delays and failures when metadata do not live up to expectations. But more importantly, the community must acknowledge the lack of established standards, methodologies or tools for metadata quality evaluation. Or to put it in the often-cited words of Diane Hillmann: "There are no metadata police".

In the absence of such standards or tools metadata practitioners usually believe that documenting the quality of their metadata is too costly a project to be undertaken. This paper shows that useful metadata indicators can be produced at a very low cost from existing metadata using general-purpose data-profiling tools. In order to facilitate the measurement and improvement of metadata we propose to integrate such tools with collection management applications, making quality measurement a continuous and seamless task. This will remove the barriers that currently prevent practitioners from actually acting on issues of metadata quality.

2. Overview of the Research

2.1. Global Data Quality Research

Metadata quality is, obviously, not only an issue for the cultural heritage sector. A large body of research, development and tools has been developed throughout the 1990's within the computer science field, the corporate world and public administrations to examine the notion of data or information quality. A multitude of other denominators and sub activities, such as data cleaning, -profiling and –standardization exist. An overview of the data quality field can be found in "Data quality: concepts, methodologies and techniques" by Batini and Scannapieco (2006) and "Data Quality : the Field Guide (2001) by Thomas Redman.

Within this large domain it is the specific topic of data profiling that is of special interest to us. Data profiling is the first and the essential step towards data quality in the sense that it consists of gathering factual information on the data quality that can be used, firstly, to decide which actions to take in order to enhance quality and, secondly, to inform users about the quality of the data they are consulting. An automated implementation of a data profiling procedure could reduce uncertainty and misconceptions regarding the quality of our collection registration databases. Collection managers and the public alike sorely need concise reports consisting of up-to-date statistical information on the quality of the totality of the records.

The application and utility of such a tool can be demonstrated by taking a look at another domain. An interesting application that might inspire methodologies and tools for the cultural heritage sector is offered by the research community around biodiversity data. The aggregation of huge sets of scientific data concerning climate, flora and fauna resulted in the same problems mentioned above. The Reference Center for Environmental Information of Brazil therefore has developed a data cleaning tool which aims to help curators identify possible errors. The system presents "suspect" records, recommending that they be checked by the author or curator.

FIG. 1: Screenshot of a data cleaning tool from the biodiversity domain (http://splink.cria.org.br/dc/)

Figure 1 represents information that is generated on the fly on the actual data by pointing out how many records are online, how many of them are geo-referenced, how many duplicated records have been detected, when the last update of the collection took place, etc. Each time suspect records are mentioned a direct link is provided to verify manually in detail the record and its metadata. Among the options offered on this page we especially would like to point out the possibility to visualize the data cleaning statistics as graphs representing the evolution through time of the number of suspect authors, duplicated records and catalog numbers (see FIG. 2).

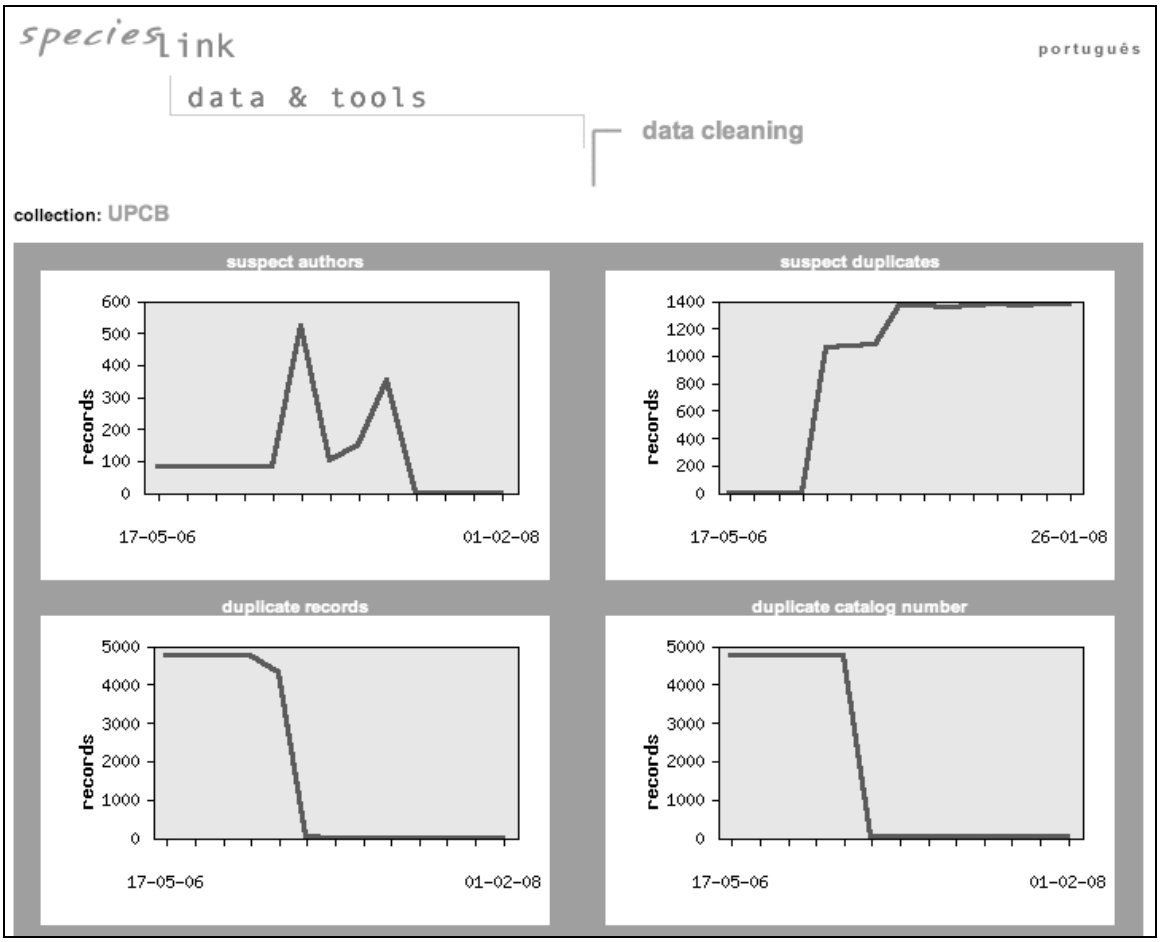

FIG. 2: Graphs representing the evolution of data quality within the biodiversity domain (http://splink.cria.org.br/dc/)

This tool offers the opportunity for a potential user of the collection to grasp within ten or fifteen minutes the quality of the data he or she is interested in.

2.2. Specificity of the Cultural Heritage Sector

Now that we have given an example from another application domain we should try to define the specific problems and characteristics related to the cultural heritage sector in order to see how tools from other domains could be applied to museum metadata.

Firstly, in contrast with information systems from other domains, such as the financial or the administrative sectors, the direct economic value of the metadata from the cultural heritage sector are comparatively limited. Metadata could play a crucial role in the re-use and marketing of digital cultural heritage, but European reports and projects investigating business models based on the commercialization of digital cultural heritage from the public domain do not point to viable options. Put simply, one cannot expect a traditional return on investment of digitization projects in the sense that the market validation of digital cultural heritage is not likely to make up for the investments made for the digitization. But this does not mean the sector cannot learn something from more economically viable domains, where data-profiling tools offer a means to introduce more accountability through statistical monitoring. The public financing of long-term metadata creation projects is unfortunately sometimes regarded as throwing money into a black hole. Data profiling could help to quantify the efficiency and effectiveness of metadata creation throughout the project life-cycle.

Secondly, museums and other heritage institutions often find it hard to define the exact needs of their users, especially when the collections consist of art. Compared to other application domains, user needs regarding cultural heritage are mostly defined in very general and vague terms. This makes metadata evaluation difficult since quality is at its most abstract level defined as the "fitness for purpose". But how can this be judged without sufficient knowledge of user expectations? Log files of user queries are haphazardly used for research purposes (Cunningham and Mahoui, 2000), but the logs have little real impact on collection description. Recent experiments with user-generated metadata such as user comments and folksonomies offer an interesting step in this direction (van Hooland, 2005). In a broad sense logs of user queries, comments and tags could also be considered as metadata linked to the collection, to which data profiling can be applied in order to more easily detect patterns and recurrences.

Lastly, we must to point out the empirical and non-structured character of cultural heritage documents. It is the core-business of heritage holders to manage and facilitate access to historical collections, for which it can be very time-consuming and sometimes impossible to document the origin and intent of the collection. Sometimes old documentation can exist, but the updating of legacy metadata is a necessity. This illustrates the problem of the ever-extendibility of metadata, in the sense that metadata themselves have to be documented as the reality and its definition evolve throughout time. But administrative or legislative institutions, which are obliged to retain their historical data, are also confronted with shifting definitions, domains and attributes of concepts such as, for example, unemployment, nationality or retirement (Boydens, 2001). The unstructured character of cultural heritage information is also blamed for the difficulty of inserting documentation into rigorously defined database fields. The extensive work in recent years on metadata models has attempted to structure as much as possible the documentation process by providing clear-cut definitions of metadata fields and sets of possible values (e.g with controlled vocabularies). But still, the descriptions that contain key information for users are contained in free-text fields. It is precisely the automated analyses of unstructured text which poses problems when assessing metadata quality.

2.3. Current Research within the Cultural Heritage Sector

The first discussions on metadata quality within the cultural heritage sector dealt with bibliographic control in the library world. However, the growing variety of types of resources, their metadata formats and user communities called for an enlarged scope. Bruce and Hillmann (2004) provide the first major theoretical foundation regarding metadata quality with their "systematic, domain- and method-independent discussion of quality indicators".

Defining quality measurements and metrics is essential, but they also have to be put into practice. The manual analysis of a limited sample of the complete set of metadata records has been a way to gather interesting indications (Shreeves et all, 2005). However, this manual approach has two obvious disadvantages: 1) it is too time consuming (and thus too expensive) and 2) it only offers a "photograph" of a sample of the metadata records at one specific moment in time. Therefore, we will focus only on practical semi-automated approaches that can repeatedly analyze the totality of a given metadata set.

Tennant (2004) proposes a minimal, pragmatic set of analysis functions to be applied on metadata and specifies queries to be computed such as the total number of occurrences of a certain value or patterns across records (e.g. all records with "x" in the "y" field do not have a "z" field). The application of such scripts or queries on large numbers of metadata records produces results which are difficult to grasp without the aid of visualization software. Dushay and Hillmann present a tool that can translate the results of queries upon a large collection of records into a human-readable form that allows the detecting of patterns and the extent of the problems (Dushay and Hillmann, 2003). Several researchers have also worked on metadata transformation and enrichment, especially in the context of aggregated content projects. Foulonneau and Cole (2005) report, for example, on how harvested records can be transformed to be of higher use in the context of an OAI service provider.

Automated quality assessment normally concentrates on what in French is referred to as *critique externe* in the context of the evaluation of historical sources: it focuses on the formal characteristics of metadata, and not on its actual content. The *critique interne* is left to human evaluation, since it is impossible to develop automated tools to grasp evaluation criteria such as accuracy and conformance to expectations. Ochoa and Duval (2007) however propose to translate these and the other criteria from the Bruce and Hillmann framework into equations that can be automatically applied. Still, this approach only applies to metadata of textual resources and not to other types of unstructured data such as images.

One of the most promising ideas has been formulated by Hillmann and Phipps (2007) who advocate the machine readability of application profiles. The real power of these "templates for expectation" can only be unleashed if their statements can be matched with the actual syntax and content of the metadata in an automated manner. But the automated validation of XML and RDF that wants to go further then just checking the "well-formedness" is still problematic, even though progress is being made (Brickley 2005).

3. Applying Data Profiling Techniques to Museum Metadata

Most of the research mentioned above used custom-written queries to be applied to the metadata records. This paper explicitly proposes to use a data profiler. Olson (2002) defines data profiling as "the use of analytical techniques to discover the true structure, content, and quality of a collection of data". We are interested to see which results can be obtained by using an open-source general-purpose data profiling tool, available at http://sourceforge.net/projects/dataprofiler/ that works in three steps. First, the analysis to perform on the dataset has to be set up by creating an XML profile specification file (see figure 3) in which is specified which analysis runs on which column of the dataset. Five analyses are at our disposal, which we will present with the help of examples from our test collection. In a second step, the profiler itself is launched, which will read the XML file and store the result of the profiling into a local repository and the information about the profiling execution into a catalog file. The catalog file is used to record what profile specification (.xml file) was used as a basis for profiling and to retrieve the results from the local repository. Third, the visualizer is run to view the profile execution results. These can then be exported for further analysis in other tools.

```
<runtime-analysis id="object_id.patternanalyzer" context="object_id" source="collection">
        <object class-name = "datadiscovery.analyzer.impl.PatternAnalyzer">
            <attribute name="columnName" value="objectid"/>
        </object>
</runtime-analysis>
<runtime-analysis id="medium.pattern" context="medium" source="collection">
        <object class-name = "datadiscovery.analyzer.impl.HistogramAnalyzer">
            <attribute name="columnName" value="medium"/>
        </object>
</runtime-analysis>
<runtime-analysis id="objectnumber.patternanalyzer" context="objectnumber" source="collection">
        <object class-name = "datadiscovery.analyzer.impl.PatternAnalyzer">
            <attribute name="columnName" value="objectnumber"/>
        </object>
</runtime-analysis>
<runtime-analysis id="description.lenghtanalyzer" context="description" source="collection">
        <object class-name = "datadiscovery.analyzer.impl.StringLengthAnalyzer">
            <attribute name="columnName" value="description"/>
        </object>
</runtime-analysis>
```

FIG. 3: Illustration of the XML profile specification file

We have tested the profiler on a comma-delimited export file from the ethnographic department of the Royal Museum for Central Africa consisting of 69,719 records, each record consisting of 13 fields (object id, object number, object count, date of collection, date of entry,

date of production, title, medium, dimensions, thesaurus terms, description, old region, actual region). The majority of the metadata are in French, with Dutch being used in a few cases.

The end result of the profiling process is the creation of a report which specifies for each metadata field a rigorous definition, the domain the values can belong to and the referential

integrity rules with other metadata fields. Results of the different analyses allow the analyst to discover violations against the definition, domain and referential integrity rules of each metadata field. We will now illustrate the different analyses with examples from our test collection.

3.1. NullCount Analysis

The NullCount analysis calculates the number of records where the specified column holds no value. Table 1 illustrates the high number of records that have no value for certain fields. Several fields, such as "description", "dimensions", "date_of_production", "date_of_collecting" and "creditline" have no value 90% of the time, which is cause for concern. Users expect values in fields, especially fields as basic as 'description.

TABLE 1: Percentage of empty fields

Fieldname	**Percentage of empty fields**
objectid	0%
objectnumber	0%
objectcount	0%
date_of_collecting	87,5%
date_of_entry	55,6%
date_of_production	92%
title	8%
medium	66.3%
dimensions	90.7%
creditline	89.5%
description	92.7%
region_old	44%
region_new	44%

3.2. Pattern Analysis

The Pattern analysis calculates the different formats used to represent values. The values can be alphabetical characters (represented by the profiler with A), numerical characters (represented by the profiler with 9) or other special signs such as a punctuation sign or a slash. This analysis is particularly useful to examine the values that correspond to a certain fixed syntax, such as accession numbers and dates. The accession number in the case of our data set has to correspond to the following fixed syntax: [collection code].[inscription year].[lot number].[number of the item within a lot]-[number that indicates that the item is a part of series]. When running the pattern analyzer, we can see that 92% of the values match the required syntax.

The different date fields also offer an excellent opportunity to apply the pattern analysis. There is a total number of 52 different ways to encode the date_of_collecting. This is due to the fact that other information is also saved within the field in some cases. Obviously, this practice should be avoided. Table 2 represents the 10 most frequent patterns used to represent the date when an item was acquired and clearly demonstrates the need to standardize the input of dates.

TABLE 2: the 10 most recurrent patterns for the date_of_collecting field.

Pattern	Number of occurrences	Example
(empty)	65011	
9999-9999	1564	1891-1912
9999	1105	1909
99-99/9999	574	09-10/1992
99/9999	347	01/1994
99-9999	346	08-1950
99/99/9999	312	04/08/1963
AAA 9999	90	Mai 1938
AAAAAAA-AAAA 9999	84	Janvier-mars 1999
99-99 9999	61	01-02 1993

The same conclusion can be drawn from the results of the pattern analysis when applied on the dimension field (see table 3). Measures are not standardized (both mm and cm are used) and apparently no rules were laid down regarding the syntax. As in the case of the problem with dates, this incoherence makes the searching difficult, not to say completely impossible. The output of this type of analysis can be used to develop scripts for normalization and to build up value vocabularies.

TABLE 3: examples of different patterns to describe dimensions.

Pattern	Number of occurrences	Example
99 A 99 AA	1190	13 x 18 cm
999 AA	388	920 mm
999 A 999	382	573 x100
99 AA A 99AA	196	37 mm x 16 mm
99 AA A 99 AA A 99 AA	107	52 cm x 25 cm x 25 cm
99	14	72

3.3. Histogram Analysis

The histogram analysis produces a histogram of the different values that exist for a specific metadata field. We can apply this analysis to quite a range of fields. Table 4 represents for example the titles that appear more than a thousand times throughout the collection. These data can serve as an excellent guide for discussions regarding the precision of the terms used in fields.

TABLE 4: Most frequent titles.

Title	Number of occurrences
(empty)	5623
statuette	2043
panier	1800
bracelet	1792
collier	1376
masque	1324
groupe	1250
couteau	1073
sifflet en bois	1012

"By accident" strange values may be discovered by this analysis. For example, when applied to the field "object_count" the histogram analysis shows us that 39 fields have the value "0", which is a violation of domain range integrity since an object must at least consist of one item.

3.4. Case Analysis

The case analysis gives an overview of the use of capitalized and non-capitalized alphabetic characters. The application of this analysis is rather limited but still enables one to check the level of consistency of the metadata input.

TABLE 5: Use of upper- and lowercase characters.

Case type	**Number of occurrences**	**Frequency (on the total number of non-empty fields)**
Mixed case	21186	54.7%
All uppercase	14889	38.4%
All lowercase	2645	6.8%

3.5. Length Analysis

The length analysis calculates the number of characters used in a field. Again, this is a very basic query that is performed on the metadata but its application can lead to interesting and unexpected results. When applied to the field "objectnumber", the profiler informs us that 69,718 values consist of 42 characters and one value consists of 55 characters, although we see that the format of this field varies and never takes up 42 characters. The most frequent pattern "AA.9999.99.99" only consists of 13 characters, so where do these values come from? Figure 9 shows the reason behind these values. A copy/paste of the data within a text editor such as Word reveals the formatting of the characters and explicitly shows the whitespaces that are included within each value. The same phenomenon appears for the field "date_of_production". Although the waste of storage space within the database is perhaps no longer a critical issue, the discrepancy between how the values are perceived and their true composition can poses problems for the long-term preservation of the metadata.

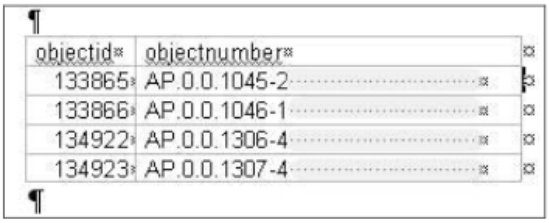

FIG 4: Presence of whitespaces within values.

4. Research and Development Agenda: Internalizing Metadata Quality within the Creation Workflow

The different analyses illustrated above clearly prove that simple and inexpensive data profiling techniques can bring many problems or particularities within large sets of metadata to the surface quite easily. But applying external tools on a periodic basis remains too much an ad-hoc solution to serve as an effective management tool for metadata quality improvement activities. And just as with manual sampling methods it only produces a "photograph" of the state

of the metadata records at a specific moment in time. Ochoa and Duval (2007) point out a soft spot when they refer to metadata quality analysis as a "research activity with no practical implications in the functionality or performance of the digital repository."

The only way to effectively have a day-to-day impact on metadata quality is to seamlessly implement a data profiling procedure within the metadata creation workflow. In the context of museum metadata the collection management system should thus incorporate functionality that enables collection managers to automatically draw data profiling reports with statistics and graphs that enable the continuous monitoring of the evolution of metadata quality.

No existing software offers such functionality. Therefore, we have established a collaboration with the development team of the open-source collection management software OpenCollection to develop and implement a metadata quality tool within that software package. OpenCollection is a general-purpose collection management system intended for use with a wide variety of materials. Current users include representatives from many fields, including fine art, anthropology, film, oral history, local history, architecture, material culture, biodiversity conservation, libraries, corporate archives and digital asset management. The most important features concerning metadata management are :

1. Completely web-based user interface, meaning that metadata input can be very easily distributed among a large group of indexers/catalogers or external experts.
2. Configurable, type-specific user defined key/value attribute system. In addition to the standard set of OpenCollection fields representing concepts applicable to anything that can be cataloged — things like "accession number" — sets of attributes functioning as repeatable custom fields,) may be defined. These sets can map to established metadata standards such as Dublin Core, Darwin Core, VRA Core 3.0, CDWA Lite, et. al. Attribute sets may be type-specific: they can be defined such that they are only available for specific types of cataloged items (ex. photographs, video tapes, films). They may also be repeating, and it is possible to impose an intrinsic data type (text, integer or floating point number, date) as well as bounds and pattern-based input validation rules.
3. Automatic extraction of metadata from uploaded media files.
4. Extensive support for authority lists and controlled vocabularies. A tool is included to import Getty Art and Architecture Thesaurus (AAT) data files.

We are currently evaluating several strategies for integration of the metadata quality tools described in this paper with OpenCollection. These range from straightforward inclusion of metrics generated by our tool in OpenCollection's reporting system to more interactive approaches built into the metadata creation workflow itself. Examples of the latter include:

1. Dynamic evaluation during input of attributes, with display of quality/suitability metrics and, when possible, suggestions for improvement.
2. Visible per-record and per-field indicators of measured quality. The indicators are color coded and can provide detailed quality metrics on-demand.
3. Expansion of the OpenCollection search engine to support searches on quality metrics. Metric search criteria may be freely mixed with traditional content-based search terms, enabling users to efficiently locate groups of related problematic data.

The seamlessly integrated metadata quality module would be packaged with analyses available out-of-the-box. This would allow metadata practitioners to have a clear view on the state of their metadata. Hopefully, getting this first "general" summary for free will catch their attention to the metadata quality issue and drive them to improve quality.

5. Conclusions

This article has given a concise overview of the metadata quality issue and its specific nature within the cultural heritage sector. Secondly, a general-purpose data-profiling tool has been applied to a large test-collection of museum metadata which resulted in the identification of various problems and particularities in the metadata. Taking these results a step further we are finally promoting a pro-active way of dealing with metadata quality by endeavoring to directly incorporate a methodology and tool in an open-source collection management system. This innovative approach will introduce more accountability into the metadata creation process as a whole, which is at the moment all too often considered as a form of black art.

Acknowledgements

The authors would like to thank Hein Vanhee of the Royal Museum for Central Africa for supplying the test-data and the students of the course "Qualité de l'information et des documents numériques" of Professor Isabelle Boydens and Yves Bontemps of the Information and Communication Science Department of the Université Libre de Bruxelles, for performing analyses on the test-data.

References

Batini, Carlo, and Monica Scannapieco. (2006). *Data quality: Concepts, methodologies and techniques*. New York: Springer.

Boydens, Isabelle. (2001). *Informatique, normes et temps*. Bruxelles: Bruylandt.

Brickley, Dan. (2005). *CheckRDFSyntax and Schemarama Revisited.* Retrieved May 20, 2008, from http://danbri.org/words/2005/07/30/114.

Bruce, Thomas, and Diane Hillmann. (2004). The continuum of metadata quality: Defining, expressing, exploiting. In Diane I. Hillmann & Elaine L. Westbrooks (Eds.), *Metadata in practice,* (pp. 238-256). Chicago: American Library Association.

Cunningham, Sally Jo, and Malika Mahoui. (2000). A comparative transaction log analysis of two computing collections. *4th European Conference on Digital Libraries,* (pp. 418-423).

Dushay, Naomi, and Diane Hillmann. (2003). Analyzing metadata for effective use and re-use. *DC-2003 Dublin Core Conference: Supporting Communities of Discourse and Practice - Metadata Research and Applications, September 28 - October 3, 2003.* Retrieved February 14, 2008, from http://www.siderean.com/dc2003/501_Paper24.pdf.

Hillmann, Diane, and Jon Phipps. (2007). Application profiles: Exposing and enforcing metadata quality. *Proceedings of the International Conference on Dublin Core and Metadata Applications*, *Singapore.* Retrieved May 20, 2008, from http://www.dcmipubs.org/ojs/index.php/pubs/article/viewFile/41/20.

Foulonneau, Muriel, and Timothy Cole. (2005). Strategies for reprocessing aggregated metadata. *Lecture notes in computer science 3652,* (pp. 290-301). Berlin, Heidelberg: Springer, ECDL. Retrieved February 14, 2008, from http://cicharvest.grainger.uiuc.edu/documents/metadatareprocessing.pdf.

Ochoa, Xavier, and Erik Duval. (2007). Towards automatic evaluation of metadata quality in digital repositories. *Ariadne.* Retrieved February 14, 2008, from http://ariadne.cti.espol.edu.ec/M4M/files/TowardsAutomaticQuality.pdf.

Olson, Jack. (2002). *Data quality: The accuracy dimension*. San Francisco: Morgan Kaufman.

Redman, Thomas. (2001). *Data quality:The field guide*. New Jersey, Boston: Digital Press.

Shreeves, Sarah, Ellen Knutson, Besiki Stvilia, Carole Palmer, Michael Twidale, and Timothy Cole. (2005). Is « quality » metadata « shareable » metadata ? The implications of local metadata practices for federated collections. *ACRL Twelfth National Conference*. Minneapolis: ALA.

Tennant, Roy. (2004). *Specifications for metadata processing tools.* 2007, 1(2), California Digital Library. Retrieved February 14, 2008, from http://www.cdlib.org/inside/projects/harvesting/metadata_tools.htm.

Van Hooland, Seth. (2005). Spectator becomes annotator: Possibilities offered by user-generated metadata for image databases. *Proceedings CILIP Cataloguing & Indexing Group Annual Conference, University of East Anglia, UK, 13-15 September 2006.* Retrieved February 14, 2008, from http://homepages.ulb.ac.be/~svhoolan/Usergeneratedmetadata.pdf.