Available online at www.sciencedirect.com
ScienceDirect
Information Processing and Management 43 (2007) 1416–1437
INFORMATION PROCESSING & MANAGEMENT
www.elsevier.com/locate/infoproman

"What is a good digital library?" – A quality model for digital libraries

Marcos André Gonçalves [a,*], Bárbara L. Moreira [a], Edward A. Fox [b], Layne T. Watson [b]

[a] Department of Computer Science, Federal University of Minas Gerais, 31270-901 Belo Horizonte, MG, Brazil
[b] Department of Computer Science, Virginia Tech, Blacksburg, VA 24061, USA

Received 25 April 2006; received in revised form 22 November 2006; accepted 22 November 2006
Available online 6 March 2007

Abstract

In this article, we elaborate on the meaning of quality in digital libraries (DLs) by proposing a model that is deeply grounded in a formal framework for digital libraries: 5S (Streams, Structures, Spaces, Scenarios, and Societies). For each major DL concept in the framework we formally define a number of dimensions of quality and propose a set of numerical indicators for those quality dimensions. In particular, we consider key concepts of a minimal DL: catalog, collection, digital object, metadata specification, repository, and services. Regarding quality dimensions, we consider: accessibility, accuracy, completeness, composability, conformance, consistency, effectiveness, efficiency, extensibility, pertinence, preservability, relevance, reliability, reusability, significance, similarity, and timeliness. Regarding measurement, we consider characteristics like: response time (with regard to efficiency), cost of migration (with respect to preservability), and number of service failures (to assess reliability). For some key DL concepts, the (quality dimension, numerical indicator) pairs are illustrated through their application to a number of "real-world" digital libraries. We also discuss connections between the proposed dimensions of DL quality and an expanded version of a workshop's consensus view of the life cycle of information in digital libraries. Such connections can be used to determine when and where quality issues can be measured, assessed, and improved – as well as how possible quality problems can be prevented, detected, and eliminated.

Keywords: Quality; Evaluation; Digital libraries; Theory; Information life cycle

1. Introduction

What is a good digital library? As was pointed out in Fuhr, Hansen, Mabe, Micsik, and Sölvberg (2001), the answer to this question depends on whom you ask. Many consider that what differentiates a good DL from a not so good one is the quality of its services and content. In previous work, we have sought to formally elaborate the notion of digital libraries using the 5S framework (Gonçalves, Fox, Watson, & Kipp, 2004).

* Corresponding author. Tel.: +55 31 3499 5000; fax: +55 31 3499 4188.
E-mail address: mgoncalv@dcc.ufmg.br (M.A. Gonçalves).

0306-4573/$ - see front matter
doi:10.1016/j.ipm.2006.11.010

Since one of the main goals of that work with 5S was to try to answer (at least partially) the question "What is a digital library?" our hypothesis in this article is that further development of the theory will allow us to define critical dimensions and indicators of DL quality. The "digital" nature of digital libraries allows automatic assessment and enforcement of those quality properties, thereby supporting prevention and elimination of quality problems, which may be more difficult in physical libraries. 5S gives a standard terminology to discuss these issues in a common framework. Moreover, the formal nature of our DL theory allows us to add precision as we define specific DL quality dimensions and corresponding numeric indicators.

In this article, we will follow the standard terminology used in the *social sciences* (Babbie, 1990). We will use the term *composite quality indicator*[1] (or in short *quality indicator*) to refer to the proposed quantities instead of the stronger term *quality measure*. Only after one has a number of indicators, and they are validated[2] and tested for reliability,[3] can they be composed into reliable "measures". Despite partial tests of validity (for example, through focus groups)[4] the proposed quality indicators do not qualify as measures yet. Also, it should be stressed that the proposed quantities are only approximations of or give quantified indication of a quality dimension. They should not be interpreted as a complete specification of a quality dimension, since more factors/variables could be relevant than are specified here. We will, however, reserve the right to use the term "measure" when talking about standard measures that have long been used by the CS/LIS communities. The distinction should be clear in context.

This article is organized as follows. Section 2 provides background and context necessary to understand the remainder of the article. Sections 3–6 present all the dimensions of quality, the proposed indicators, and their applications to key DL concepts. Section 7 deals with the connections between the proposed dimensions and Borgman et al.'s Information Life Cycle (Borgman, 1996). Section 8 shows the evaluation of the proposed quality model with a focus group. Section 9 covers related work and Section 10 concludes the article.

2. Background and context

In this section, we summarize the 5S theory from Gonçalves et al. (2004). Here we take a minimalist approach, i.e., we define, according to our analysis, the minimum set of concepts required for a system to be considered a digital library. Accordingly, let:

- *Streams* be a set of streams, which are sequences of arbitrary types (e.g., bits, characters, pixels, frames).
- *Structs* be a set of structures, which are tuples, (G, L, F), where $G = (V, E)$ is a directed graph and $F : (V \cup E) \rightarrow L$ is a labeling function.
- *Sps* be a set of spaces each of which can be a measurable, measure, probability, topological, metric, or vector space.
- $Scs = \{sc_1, sc_2, \ldots, sc_d\}$ be a set of scenarios where each $sc_k = \langle e_{1k}(\{p_{1k}\}), e_{2k}(\{p_{2k}\}), \ldots, e_{d_k k}(\{p_{d_{kk}}\})\rangle$ is a sequence of events that also can have a number of parameters p_{ik}. Events represent changes in computational states; parameters represent specific variables defining a state and their respective values.
- St^2 be a set of functions $\Psi : V \times Streams \rightarrow (N \times N)$ that associate nodes of a structure with a pair of natural numbers (a, b) corresponding to a segment of a stream.
- $Coll = \{C_1, C_2, \ldots, C_f\}$ be a set of DL collections where each DL collection $C_k = \{do_{1k}, do_{2k}, \ldots, do_{f_k}k\}$ is a set of digital objects. Each digital object $do_k = (h_k, Stm_{1k}, Stt_{2k}, \Omega_k)$ is a tuple where $Stm_{1k} \subseteq Streams$, $Stt_{2k} \subseteq Structs$, $\Omega_k \subseteq St^2$, and h_k is a handle which represents a unique identifier for the object.
- $Cat = \{DM_{C_1}, DM_{C_2}, \ldots, DM_{C_f}\}$ be a set of metadata catalogs for *Coll* where each metadata catalog $DM_{C_k} = \{(h, mss_{hk})\}$, and $mss_{hk} = \{ms_{hk1}, ms_{hk2}, \ldots, ms_{hkn_{hk}}\}$ is a set of descriptive metadata specifications.

[1] An indicator composed of two or more simpler indicators or variables.

[2] According to Babbie (1990), validity refers to the extent to which a specific measurement provides data that relate to commonly accepted meanings of a particular concept. There are numerous yardsticks for determining validity: face validity, criterion-validity, content validity, and construct validity.

[3] Also according to Babbie (1990), reliability refers to the likelihood that a given measurement procedure will yield the same description of a given phenomena if that measurement is repeated.

[4] A type of face validity.

Table 1
DL high-level concepts and corresponding DL dimensions of quality with respective metrics

DL concept	Dimension of quality	Factors/variables involved in measuring
Digital object	Accessibility	Collection, # of structured streams, rights management metadata, communities
	Pertinence	Context, information, information need
	Preservability	Fidelity (lossiness), migration cost, digital object complexity, stream formats
	Relevance	Query (representation), digital object (representation), external judgment
	Similarity	Same as in relevance, citation/link patterns
	Significance	Citation/link patterns
	Timeliness	Age, time of latest citation, collection freshness
Metadata specification	Accuracy	Accurate attributes, # of attributes in the record
	Completeness	Missing attributes, schema size
	Conformance	Conformant attributes, schema size
Collection	Completeness	Collection size, size of the 'ideal collection'
Catalog	Completeness	# of digital objects without a set of metadata specifications, size of the described collection
	Consistency	# of sets of metadata specifications per digital object
Repository	Completeness	# of collections
	Consistency	# of collections in repository, catalog/collection pair-wise consistency
Services	Composability	Extensibility, reusability
	Efficiency	Response time
	Effectiveness	Precision/recall (search), F1 measure (classification)
	Extensibility	# of extended services, # of services in the DL, # of lines of code per service manager
	Reusability	# of reused services, # of services in the DL, # of lines of code per service manager
	Reliability	# of service failures, # of accesses

Each descriptive metadata specification ms_{hki} is a structure with atomic values (e.g., numbers, dates, strings) associated with nodes.

- Repository $R = (\{C_i\}_{i=1}^{f}, \{get, store, delete\})$ be a set of collections along with operations to manipulate them (see Gonçalves et al. (2004) for details on the semantics of these operations).
- $Serv = \{Se_1, Se_2, \ldots, Se_s\}$ be a set of services where each service $Se_k = \{sc_{1k}, \ldots, sc_{s_k}k\}$ is described by a set of related scenarios. Any digital library should contain at least services for browsing, indexing, and searching.
- $Soc = (Comm, S)$ where $Comm$ is a set of communities and S is a set of relationships among communities. $SM = \{sm_1, sm_2, \ldots, sm_j\}$ and $Ac = \{ac_1, ac_2, \ldots, ac_r\}$ are two such communities, where the former is a set of service managers responsible for running DL services and the latter is a set of actors that use those services.[5] Being basically an electronic entity, a member sm_k of SM distinguishes itself from actors by defining or implementing a set of operations $\{op_{1k}, op_{2k}, \ldots, op_{n_k}k\} \subset sm_k$. Each operation op_{ik} of sm_k is characterized by a triple $(n_{ik}, sig_{ik}, imp_{ik})$, where n_{ik} is the operation's name, sig_{ik} is the operation's signature (which includes the operation's input and output parameters), and imp_{ik} is the operation's implementation. These operations define the capabilities of a service manager sm_k. For example, SearchManager ∍ match(q:query, C:collection)[6] indicates that a SearchManager contains or defines an operation "match" with two parameters: a query and a collection.

According to the 5S formal framework a digital library is a 4-tuple (R, Cat, $Serv$, Soc).

The set of main DL concepts defined above was derived from a comprehensive review of the DL related literature as described in Gonçalves et al. (2004). We started from these higher level concepts defined above, and then from a similar process of comprehensive literature review on quality issues in information systems in

[5] In this paper we will focus only on the relationships between and among actors and service managers that correspond to interactions mediated by the DL. We will not focus on interactions that happen outside of the system, such as social interactions among human actors.

[6] To simplify notation, we will represent an operation $op_x = (n_x, sig_x, imp_x)$ by $n_x(\{p_{xk}\})$, where $\{p_{xk}\}$ is the set of input parameters of op_x. The output parameters and implementation can be added when a fuller description of the operation is required. Note that our focus on the input parameters is in conformance with specifying the context for an operation.

Table 2
Dimensions of quality and Ss involved in their definitions

DL concept	Dimension of quality	Some 'S' concepts involved
Digital object	Accessibility	Societies (actor), Structures (metadata specification), Streams + Structures (structured streams)
	Pertinence	Societies (actor), Scenarios (task)
	Preservability	Streams, Structures (structural metadata), Scenarios (process (e.g., migration))
	Relevance	Streams + Structures (structured streams), Structures (query), Spaces (Metric, Probabilistic, Vector)
	Similarity	Same as in relevance, Structures (citation/link patterns)
	Significance	Structures (citation/link patterns)
	Timeliness	Streams (time), Structures (citation/link patterns)
Metadata specification	Accuracy	Structure (properties, values)
	Completeness	Structure (properties, schema)
	Conformance	Structure (properties, schema)
Collection	Completeness	Structure (collection)
Catalog	Completeness	Structure (collection)
	Consistency	Structure (collection)
Repository	Completeness	Structure (collection)
	Consistency	Structure (catalog, collection)
Services	Composability	See Extensibility, reusability
	Efficiency	Streams (time), Spaces (operations, contraints)
	Effectiveness	See Pertinence, Relevance
	Extensibility	Societies + Scenarios (extends, inherits_from, redefines)
	Reusability	Societies + Scenarios (includes, reuses)
	Reliability	Societies + Scenarios (uses, executes, invokes)

general, as well as from our own experience in building DLs since 1991, we derived the list of quality dimensions described below.

Table 1 shows a summary of proposed candidate dimensions of quality for some of the most important DL concepts defined above and factors affecting the measurement of the corresponding quality dimensions.[7] The following sections explain these indicators in detail by:

(1) motivating them and discussing their meaning/utilization;
(2) formally defining them and specifying their corresponding numerical computation; and
(3) illustrating their use by applying the indicators/metrics in the context of some "real-world" DLs (e.g., ACM DL, CITIDEL, NDLTD).

Table 2 connects the proposed dimensions with some 'S'-related concepts involved in their definition. In the same way that the formalized 5S theory helps to precisely define the higher level DL concepts used here, we will use these formalizations to help define the quality indicators and their corresponding computations.

3. Digital objects

3.1. Accessibility

A digital object is accessible by a DL actor or patron, if it exists in the repository of the DL, a service is able to retrieve the object, and: (1) an overly restrictive rights management property of a metadata specification does not exist for that object; or (2) if such exists, the property does not restrict access for the particular society

[7] For simplicity, we focus on a DL concept and an indicator, not mentioning all other DL concepts that also relate. Thus, while we assign "relevance" to "digital" object, we are aware that users and queries clearly are involved too.

to which the actor belongs or to that actor in particular. A quality indicator for calculating accessibility is a function, which depends on all those factors and the granularity of the rules (e.g., entire object, structured streams). It should be noted that digital object accessibility as defined here is different from the common view of ''Web site accessibility'', which is concerned with creating better ways to provide Web content to users with disabilities (Rowan, Gregor, Sloan, & Booth, 2000). For reasons of space we omit discussion of indicators associated with that type of accessibility.

Let *access constraint* be a property of some metadata specification of a digital object do_x whose values include the sets of communities that have the right to access specific (structured) streams within the object. Also let $struct\_streams(do_x) = \Omega_x$ be the set of structured streams of do_x. The accessibility $acc(do_x, ac_y)$ of a digital object do_x to an actor ac_y is:

- 0, if there is no collection C in the DL repository R such that $do_x \in C$;
- otherwise $acc = \left(\sum_{z \in struct\_streams(do_x)} r_z(ac_y)\right) \Big/ |struct\_streams(do_x)|$, where $r_z(ac_y)$ is a rights management rule defined as an indicator function:
  1, if
    z has no access constraints; or
    z has access constraints and $ac_y \in cm_z$, where $cm_z \in Soc(1)$ is a community that has the right to access z; and
  0, otherwise.

Notice that, from a broader perspective, the accessibility of a given digital object could be affected, not only by rights management, but also by technological constraints, such as lack of Acrobat Reader to open a full-text paper in PDF format, temporary network disconnection, or restriction on the number of simultaneous users, etc. In this work, though, we have tried to focus on easily measurable intrinsic properties of the objects themselves and of the relationships between the actor and the objects.

Example of use. At Virginia Tech, a student can choose, at the moment of submission, to allow her electronic thesis or dissertation (ETD) to be viewed worldwide, by only those at the originating university, or not at all. The ''mixed'' case occurs when some portions (e.g., particular chapters or multimedia files) have restricted access while others are more widely available. The majority of Virginia Tech students choose their documents to be viewable worldwide; some initially choose not to grant worldwide access, because of concerns regarding patents or publication of results in journals/conferences.

Therefore the accessibility $acc(etd_x, ac_y)$ of a Virginia Tech ETD etd_x is:

- 0, if etd_x does not belong to the VT-ETD collection;
- otherwise $\left(\sum_{z \in struct\_streams(etd_x)} r_z(ac_y)\right) \Big/ |struct\_streams(etd_x)|$, where $r_z(ac_y)$ is a rights management rule defined as an indicator function:
  1, if
    etd_x is marked as ''worldwide access'' or
    etd_x is marked as ''VT only'' and $ac_y \in VT_{cmm}$, where VT_{cmm} is the community of Virginia Tech ID holders accessing z through a computer with a Virginia Tech registered IP address.
  0, otherwise.

Table 3 shows a partial view of the numbers of VT-ETDs that are unrestricted (worldwide, accessibility = 1 to everybody), restricted to the VT campus (accessibility = 0 worldwide, 1 to members of VT_{cmm}), or mixed,

Table 3
Accessibility of VT-ETDs (first column corresponds to the first letter of author's name)

First letter of author's name	Unrestricted	Restricted	Mixed	Degree of accessibility for users not in the VT community
A	164	50	5	mix(0.5, 0.5, 0.167, 0.1875, 0.6)
B	286	102	3	mix(0.5, 0.5, 0.13)
C	231	108	7	mix(0.11, 0.5, 0.5, 0.5, 0.33, 0.09, 0.33)
D	159	54	2	mix(0.875, 0.666)
E	67	26	1	mix(0.5)

along with the degree of accessibility $acc(etd_x, ac_y)$ of the mixed ETDs for non-VT_{cmm} members ac_y, as of March 25, 2003. For example, five out of the six chapters (structured streams) of the third mixed ETD under the letter A were available only to VT. The rights management rule therefore is 0 for all those chapters, thus making its overall accessibility to non-VT actors 1/6 or 0.167. Note that accessibility for the Virginia Tech ETDs has improved since 2003; indicators like this may be of help for those who work on collection development policies.

3.2. Pertinence

Pertinence is one of the most ‘‘social’’ quality indicators since it is a relation between the information carried by a digital object and an actor's information need. It depends heavily on the actor's knowledge, background, current task, etc.

Let $Inf(do_i)$ represent the ‘‘information’’[8] (not physical) carried by a digital object do_i in any of its components, $IN(ac_j)$ be the information need[9] of an actor ac_j, and $Context(ac_j, k)$ be an amalgam of societal factors that affect the judgment of pertinence of do_i by ac_j at time k. These include, among others, task, time, place, the actor's history of interaction, and a range of other factors that are not given explicitly but are implicit in the interaction and ambient environment. A complete formalization of context is out of the scope of this work. The reader is referred to a workshop on ‘‘Context in Information Retrieval’’ for a number of papers on the subject (Ingwersen, van Rijsbergen, & Belkin, 2004).

Also, we define, for future reference, two time dependent sub-communities of actors, *users* and *external-judges* $\subset Ac$, as:

- *users*: set of actors with an information need who use DL services to try to fulfill/satisfy that need,
- *external-judges*: set of actors responsible for determining the relevance (see Section 3.4) of a document to a query. We also assume that an external-judge cannot be assigned to judge the relevance of a document to a query representing her own information need, i.e., at each point in time $users \cap external - judges = \emptyset$.

The pertinence of a digital object do_i to a user ac_j at a time k is an indicator function[10] $Pertinence(do_i, ac_j, k): Inf(do_i) \times IN(ac_j) \times Context(ac_j, k)$ defined as:

- 1, if $Inf(do_i)$ is judged by ac_j to be informative with regards to $IN(ac_j)$ in context $Context(ac_j, k)$;
- 0, otherwise.

Since pertinence is a subjective judgment made by a *user* in a particular context it can ultimately can be accessed only by the user themself.

3.3. Preservability

Preservability is a quality property of a digital object that reflects a state of the object that can vary due to changes in hardware (e.g., new recording technologies), software (e.g., release of a new version of the software used to create/display the object), representation formats (e.g., new image standard such as JPEG 2000), and processes to which the object is submitted (e.g., migration).

There are four main technical approaches to digital preservation:

(1) *Migration*: transforming from one digital format to another format, normally a successive subsequent one (e.g., from JPEG to JPEG 2000) (Crespo & Garcia-Molina, 1998).

[8] Information and information need, by themselves, are hard notions to formally define. One comprehensive attempt is presented in Mizzaro (1996).

[9] Certain authors such as Taylor (1968) and Mizzaro (1998) make a distinction between the ‘‘real’’ and the ‘‘perceived’’ information need. We will not make this distinction here, in the interest of brevity.

[10] We agree with Voorhees (2001), Greisdorf (2003), and others who argue for non-binary pertinence/relevance functions, but such is not normal practice. We will leave extensions to our definitions for these cases for future work.

(2) *Emulation*: re-creating the original operating environment by saving the original programs and or creating new programs that can emulate the old environment (Rothenberg, 2000).
(3) *Wrapping*: packaging the object to be preserved with enough human readable metadata to allow it to be decoded in the future (Waugh, Wilkinson, Hills, & Dell'oro, 2000).
(4) *Refreshing*: copying the stream of bits from one location to another, whether the physical medium is the same or not (Levy, 1998).

Note that here we are not considering physical deterioration of the medium in which the object is stored, since this is a property of the medium itself, not the object. However, we acknowledge that this is an important problem, for which "refreshing" is the normally used approach.

For cost, operational, and technical reasons, migration is the most widely used of the three techniques mentioned above (Waugh et al., 2000). However, the ideal solution should be some combination of all the techniques (Hunter & Choudhury, 2003; Waugh et al., 2000). One example that applies such a combination is the UVC-based approach (Lorie, 2001). Nonetheless, for the purpose of the discussion below, we will concentrate on migration issues.

A digital object's preservability can be affected by its obsolescence and the fidelity of the migration process (see Fig. 1). Obsolescence reflects the fact that a very obsolete object is really hard and costly to migrate, given the difficulty of finding appropriate migration tools and the right expertise. Fidelity reflects the differences between the original and the migrated object or, in other words, reflects the distortion or the loss of information inherent in the migration process that is absorbed by the object. The more obsolete the object, and the less faithful the migration process, the lower the object's preservability. Preservability also is affected by contextual issues of specific DLs. For example, while it is desirable to always use the most faithful migration process, a DL custodian may not have sufficient resources (money, storage, personnel) to apply that process to its digital objects during migration. Based on the above discussion and on the fact that these two factors are orthogonal, we can define the preservability of a digital object do_i in a digital library dl as a tuple:

$$preservability(do_i, dl) = (fidelity\ of\ migrating(do_i, format_x, format_y), obsolescence(do_i, dl)). \quad (1)$$

As mentioned before, obsolescence is a complex notion to capture, depending on many contextual factors. Since the choice of how to deal with obsolescence generally depends on resources at the disposal of the DL custodian, one possible idea is to approximate its value by using the actual cost of migrating the object (Sanett, 2003). While a complete cost model for preservability/obsolescence is beyond the scope of this work, we recognize many factors that can affect the cost, including:

- capital direct costs:
  - software development/acquiring or license updating for newer versions,
  - hardware (for preservation processing or storage);

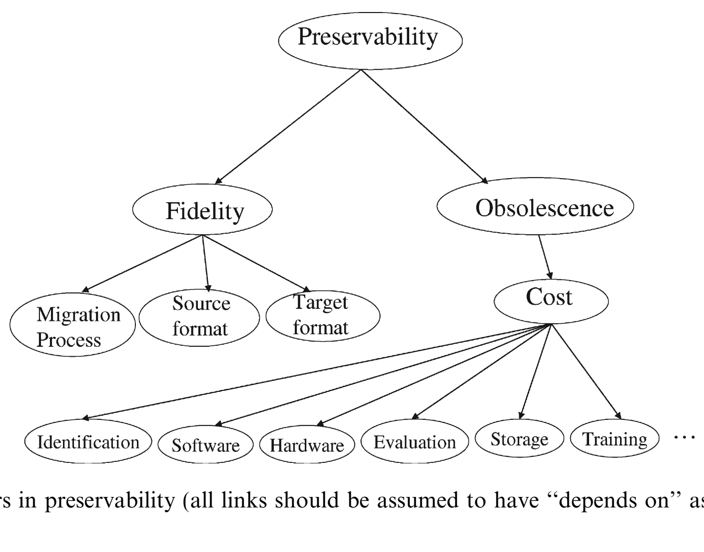


Fig. 1. Factors in preservability (all links should be assumed to have "depends on" as their labels).

- indirect operating costs:
  - identifying candidate digital objects,
  - evaluating/examining/negotiating intellectual property issues and rights,
  - storage, and
  - staff training (on software/procedures).

Obsolescence then can be defined as $obsolescence(do_i, dl) =$ cost of converting/migrating the digital object do_i within the context of the specific digital library *dl*.

The fidelity of the migration process *p* of a digital object do_i from $format_x$ to $format_y$ can be defined based on the inverse of the distortion or noise introduced by the migration process *mp*, i.e.,

$$fidelity(do_i, format_x, format_y) = \frac{1}{distortion(mp(format_x, format_y)) + 1.0}.$$

Distortion can be computed in a number of ways depending on the type of object and transformation (Sayood, 1996). One very common measure, when converting between similar formats, is the *mean squared error (mse)*. In the case of a digital object, *mse* can be defined as follows. Let $\{x_n\}$ be a stream of a digital object do_i and $\{y_n\}$ be the converted/migrated stream; the mean squared error $mse(\{x_n\},\{y_n\}) = \frac{1}{N} * \sum_{n=1}^{N}(x_n - y_n)^2$, where *N* is the size of each stream. The average mean square error for the whole object do_i can be calculated as the average of *mse* for all its streams. This assumes that the other components (graphs and functions) of a digital object will be converted exactly.

Example of Use. Let us consider the following scenario adapted from Hunter and Choudhury (2004). In 2004, a librarian receives an email notifying her that a special collection[11] of 1000 digital images, stored in TIFF version 5.0, is in danger of becoming obsolete, due to the fact that the latest version of the display software no longer supports TIFF 5.0. The librarian decides to migrate all digital photos to JPEG 2000, which now has become the de facto image preservation standard, recommended by the Research Libraries Group (RLG) (Hunter & Choudhury, 2004).

The librarian does a small search for possible migration options and finds a tool, costing $500, which converts TIFF 5.0 directly to JPEG 2000. Let us consider that the amount of time taken by the librarian and the system administrator to order the software, install it, learn it, and apply it to all digital images combined takes 20 hours. Assume also that the hourly rate in this library is $66.60 per hour per employee.[12] In order to save space, the librarian chooses to use in the migration a compression rate which produces an average *mse* of 8 per image. In this scenario, the preservability of each digital image would correspond to: preservability (image-TIFF 5.0, *dl*) = (1/9, ($500+ $66.60 * 20)/1000) = (0.11, $1.83).

3.4. Relevance

A digital object is *relevant* (Saracevic, 1975) in the context of an expression of an information need (e.g., a query) or interest (e.g., profile) and a service (e.g., searching, recommending). A role of an information satisfaction service is to provide ways to find the most relevant information for the user, which in the case of DLs is carried by digital objects and their metadata specifications.

The relevance of a digital object to a query is an indicator function $Relevance(do_i, q)$ defined as:

- 1, if do_i is judged by an *external-judge* to be relevant to *q*;
- 0, otherwise.

The most common measures for relevance estimates/predictions are based on statistical properties of the streams of the digital object and the queries. For example, in the vector space model, this relevance is

[11] Preservation of a collection, instead of a digital object, also may involve preserving all the structures (e.g., classification schemes, etc.) imposed on the collection.

[12] One thousand eight hundred is the number of hours in a work-year (37.5 h/wk * 48 wks/yr) and $110,000 the total annual cost of an employee working for this DL, based on salary, benefits, expenses.

estimated based on the distance between the vectors representing the objects and queries (as measured by the angle between them), and the components of these vectors are derived from values such as frequency of a term in a document, the frequency of the term in the collection, document size, document structure, query size, collection size, etc. Note that, in contrast to pertinence, relevance is a relation between a *representation* of a document and a *representation* of an information need (i.e., query). Also, it is supposed to be an objective, public, and social notion that can be established by a general consensus in the field, not a subjective, private judgment between the actor and her information need (Foskett, 1972; Kemp, 1974).

The distinction we have made between pertinence and relevance is derived from a view held by part of the information science community (Cosijn & Ingwersen, 2000; Foskett, 1972; Kemp, 1974; Saracevic, 1996, 1975). We have just formalized the two notions in the context of our framework. In Saracevic's work, for example, relevance, as defined by us, is called systemic or algorithmic relevance, and is a relationship between a text and a query. Pertinence, or cognitive relevance, is a relationship between the state of knowledge and cognitive information need of a user and the objects retrieved. Cognitive correspondence, informativeness, novelty, information quality, and the like are criteria by which cognitive relevance is inferred.

The external judges should evaluate the relevance of the object to the query without the cognitive load resulting from contextual interference, therefore their judgments should be more objective and more generally applicable.

3.5. Significance

Significance of a digital object can be viewed from two perspectives: (1) relative to its pertinence or relevance or (2) in absolute terms, irrespective of particular user requirements. Absolute significance can be calculated based on *raw citedness* – the number of times a document do_i is cited, or the frequency of occurrence of citations whose target is do_i. Other factors may play a role in the significance of a document such as the prestige of the journal publishing the work, its use in courses, awards given, etc., but these are very hard to quantify/measure.

Example of Use. We used 98,000 documents from the ACM Digital Library, which corresponded to approximately 1,093,700 (outgoing) citations (average of 11.53 citations per document). Table 4 shows the top five documents in the ACM collection with the highest values of significance.

Notice that significance, as defined, is supposed to increase with time, as more people take notice of the work and acknowledge it through citations. As such, publication date affects this indicator (and timeliness, see below, as well) since older publications have more chance of being cited.

3.6. Similarity

Similarity metrics reflect the relatedness between two or more digital objects. An object similar to another relevant or pertinent object has a good chance of also having these properties, but an object *too* similar to another supposedly different object can reveal a lack of quality (e.g., plagiarism, which might be found through plagiarism software) unless it is a variant version which can be identified through a de-duping process.

Table 4
Documents with highest degree of significance

Document	Publication	Year	Significance
Computer programming as art	CACM	1974	279
A generalized processor sharing approach to flow control in integrated services networks: the single-node case	IEEE/ACM Transactions on Networking (TON)	1993	138
The entity–relationship model – toward a unified view of data	ACM Transactions on Database Systems	1976	130
A relational model of data for large shared data banks	CACM	1970	121
Revised report on the algorithmic language scheme	ACM SIGPLAN Notices	1986	116

Similarity can be measured based on the digital object's content (streams) (e.g., use and frequency of words), the digital object's internal organization (structures), or citations/linking patterns. For example, similarity between two documents can be calculated using the cosine distance between the vectors representing the documents (Baeza-Yates & Ribeiro-Neto, 1999). This idea can be expanded to calculate similarity between corresponding structured streams of documents (e.g., using their title and abstract texts). Other measures, such as "bag-of-words" and Okapi can be used to calculate similarity as well (Baeza-Yates & Ribeiro-Neto, 1999).

Similarity measures also may use link or citation information to compute the relatedness of two objects. Among the most popular citation-based measures of similarity are: co-citation (Small, 1973), bibliographic coupling (Kessler, 1963), and the Amsler measure. The last one is a combination of the previous two, so we will explain only the Amsler measure (Amsler, 1972).

According to Amsler, two documents d_i and d_j are related if (1) d_i and d_j are cited by the same document, (2) d_i and d_j cite the same document, or (3) d_i cites a third document d_k that cites d_j. Thus, let Pd_i be the set of parents of d_i, and let Cd_i be the set of children of d_i. The Amsler similarity between two pages d_i and d_j is defined as

$$Amsler(d_i, d_j) = \frac{|(Pd_i \cup Cd_i) \cap (Pd_j \cup Cd_j)|}{\max(Pd_i \cup Cd_i|, |Pd_j \cup Cd_j|)}. \quad (2)$$

Eq. (2) tells us that, the more links (either parents or children) d_i and d_j have in common, the more they are related. The absolute Amsler degree of a document d_i in collection C is defined as $\sum_{d_j \in C-\{d_i\}} Amsler(d_i, d_j)$.

Example of use. Table 5 shows the top five documents in the ACM collection we studied with the highest absolute values of Amsler.

3.7. Timeliness

Timeliness of a digital object is the extent to which it is sufficiently up-to-date for the task at hand (Pipino, Lee, & Wang, 2002). It can be a function of the time when the digital object was created, stored, accessed, or cited.

Since the timeliness of an object is directly related to the information it carries, which can still be timely even if the object is "old", a good quality indicator of this quality dimension is the time of the latest citation, since it's a measure that:

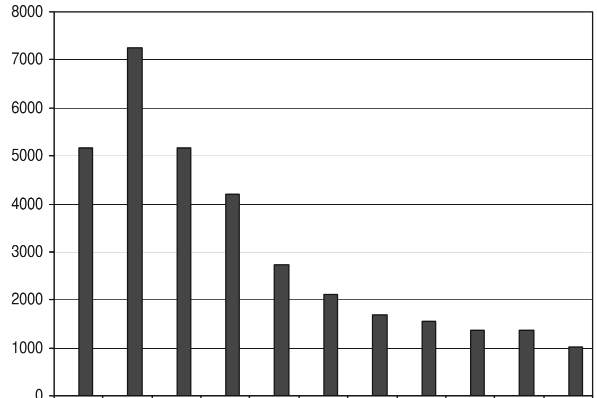
(1) captures the fact that the information carried by the object is still relevant by the time the citing object was published;
(2) is independent from the actor that receives the object and the time the object is delivered; and
(3) reflects the overall importance of the object inside its community of interest.

As it is known that many documents are never cited, an alternative is to consider the age of the object itself. Therefore the timeliness of a digital object do_i can be defined as:

- (current time or time of last freshening) – (time of the latest citation), if object is ever cited, otherwise as
- (current time or time of last freshening) – (creation time or publication time), if object is never cited.

Table 5
Documents with highest absolute Amsler degree

Document	Publication	Year	Amsler
Computer programming as an art	CACM	1974	69.15
Compiler transformations for high-performance computing	CSUR	1994	64.31
Analysis of pointers and structures	Prog. language design and implementation	1990	62.56
Query evaluation techniques for large databases	CSUR	1993	59.81
A schema for interprocedural modification side-effect analysis with pointer aliasing	TOPLAS	2001	57.90



	1	2	3	4	5	6	7	8	9	10	11
Timeliness	0	1	2	3	4	5	6	7	8	9	10
No. of Documents	5165	7264	5162	4209	2716	2120	1698	1554	1372	1357	1019

Fig. 2. Timeliness in the ACM digital library (for CITIDEL data from 2002).

Time of last freshening, which is defined as the time of the creation/publication of the most recent object in the collection to which do_i belongs, may be used instead of current time if the collection is not updated frequently. Notice that for this indicator, the lower the timeliness value, the better.

Example of use. Fig. 2 shows the distribution of timeliness (0 through 10) for documents in the ACM DL with citations. Time of last freshness is 2002. It can be seen, discounting the first set of values (timeliness = 0), that there is an inverse relation between timeliness and the size of the set of documents with that value: the smaller the value, the bigger the set, meaning that as time passes there is less chance that a document will be cited.

4. Metadata specifications and metadata format

Three main dimensions of quality can be associated with metadata specifications and metadata formats: accuracy, completeness, and conformance.

4.1. Accuracy

Accuracy is defined in terms of properties of a metadata specification for a digital object. Accuracy of a triple (r,p,v) (i.e., (*resource, property*,[13] *value*)) refers to the nearness of the value v to some value v' in the attribute range that is considered the correct one for the (resource, property) pair (r,p) (Redman, 1992). Notice that in 5S, a metadata specification is defined as a structure (G,L,F), G being a graph, L a set of labels, and F a labelling function associating components (i.e., nodes and vertices) of the graph with labels. In other words, a metadata specification can be seen as a labeled digraph. The triple $st = (F(v_i),F(e),F(v_j))$ is called a statement (derived from the descriptive metadata specification), meaning that the resource labeled $F(v_i)$ has property or attribute $F(e)$ with value $F(v_j)$. A metadata specification for a digital object is completely accurate with respect to a digital object if all the (derived) triples are accurate, assuming some appropriate accuracy threshold. The degree of accuracy of triple (r,p,v) can be defined as an indicator function or with specific rules for a particular schema/catalog. It is dependent on several factors, including the attribute's range of values v, intended use, etc. Examples are given below. Thus, the degree of accuracy $acc(ms_x)$ of a metadata specification ms_x can be defined as

$$acc(ms_x) = \frac{\sum_{\forall(r,p,v)\ \text{from}_{ms_x}} \text{degree of accuracy of } (r,p,v)}{\text{total number of triples } (r,p,v) \text{ from } ms_x} \tag{3}$$

[13] In this work we will use the terms 'metadata property', 'metadata attribute', and 'metadata field' interchangeably.

Example of Use. To illustrate the application of such an indicator we used OCLC's NDLTD Union Catalog. We chose OCLC's NDLTD Union Catalog because of its numerous problems regarding metadata accuracy, observed while creating a collection for filtering experiments (Zhang, Gonçalves, & Fox, 2003). For example, author information is very commonly found in the title field ("The concept of the church in the thought of Rudolf Bultmann – by Paul E. Stambach") and sometimes the abstract contains all kinds of information (see below) but not the thesis/dissertation's summary. We defined the following rules for the dc.author,[14] dc.title, and dc.abstract fields.

- Degree of accuracy of $(*, dc.title, *)$ for OCLC's NDLTD Union Catalog = 1, if dc.title does not contain author information; 0.5 otherwise. In case it is empty or null it receives a 0 (zero) value.
- Degree of accuracy of $(*, dc.abstract, *) = 1$ if the field corresponds to the thesis or dissertation's summary; 0 otherwise. The decision of whether a dc.abstract field corresponds to a summary or not was based on the size of the text and a number of heuristics. For example, (1) if dc.abstract is equal to "Thesis" or "Dissertation", it is not a summary; (2) if dc.abstract contains phrases like "Title from *" (e.g., "Title from first page of PDF file"), "Document formatted into pages", "Includes bibliographical references", "Mode of access", among others, it is not a summary.

According to these two rules the average OCLC accuracy for all its metadata records (approximately 14,000 records, in September 2003)[15] was calculated as around 0.79, assuming a maximum of 1.

4.2. Completeness

Completeness is a pervasive quality dimension that is associated with many of the DL concepts. The general notion of completeness can be defined as: (number of units of a particular concept)/(ideal number of units of that concept). This notion can be adapted or instantiated to specific DL concepts.

Completeness of metadata specifications refers to the degree to which values are present in the description, according to a metadata standard. As far as an individual property is concerned, only two situations are possible: either a value is assigned to the property in question, or not. The degree of completeness of a metadata specification ms_x can be defined as[16]

$$Completeness(ms_x) = 1 - \frac{\text{no. of missing attributes in } ms_x}{\text{total no. of attributes in the schema for } ms_x} \qquad (4)$$

Notice that the assumption here is that the more complete, the better. However, we acknowledge that there can be situations, for example, determined on purpose in accordance with local needs, in which this is not always true.

Example of application. Fig. 3 shows the average of completeness of all metadata specifications (records) in catalogs of the NDLTD Union Archive administered by OCLC as of February 23, 2004, relative to the Dublin Core metadata standard (15 attributes).

4.3. Conformance

The conformance of a metadata specification to a metadata standard/format/schema has been formally defined in Gonçalves et al. (2004). In that definition a value of an attribute is conformant to its schema if it has the data type of the attribute (e.g., string, date, number, etc.). That definition can be extended to include cardinality (i.e., considering mandatory/optional fields) and multiplicity (i.e., considering repeatable fields) issues.

[14] The author field in the Dublin Core standard.
[15] Over 250K in November 2006.
[16] According to the definition of conformance in Gonçalves (2004).

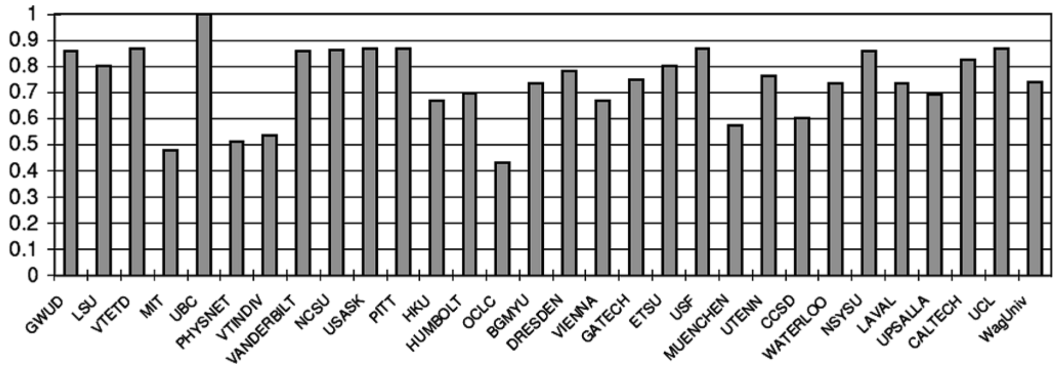


Fig. 3. Average completeness of catalogs in NDLTD (as of February 2004).

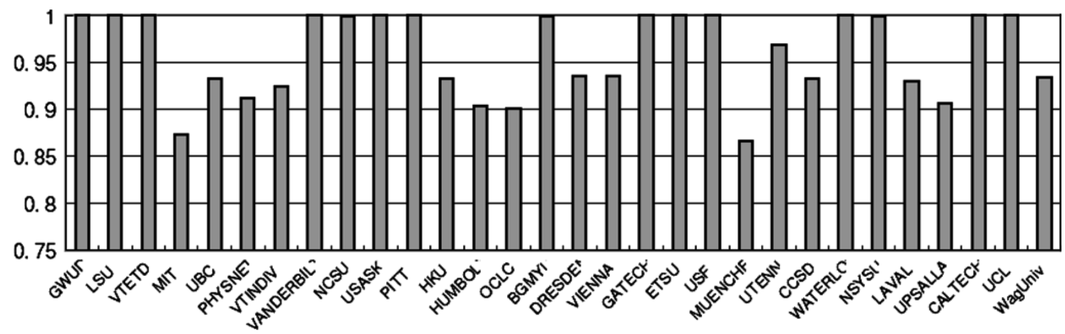


Fig. 4. Average conformance of catalogs in NDLTD (as of February 2004).

A metadata specification ms_x is *cardinally conformant* to a metadata format if:

(1) it conforms with its schema in terms of the data types of its attributes according to Definition 14 of metadata schema in Gonçalves et al. (2004);
(2) each attribute att_{xy} of ms_x appears at least once if att_{xy} is marked as mandatory in the schema; and
(3) att_{xy} does not appear more than once if it is not marked as repeatable in the schema.

From now on, we will use conformance to refer to the stronger definition of *cardinally conformant*. Different from completeness, an attribute may be missing in a metadata specification, but the attribute still can be considered conformant, if it is not marked as mandatory in the schema. The degree of conformance of a metadata specification ms_x can be defined as

$$Conformance(ms_x) = \frac{\sum_{\text{attributes } att_{xy} \text{ in schema } ms_x} \text{degree of conformance of } att_{xy}}{\text{total number of attributes in the schema for } ms_x} \tag{5}$$

The degree of conformance of att_{xy} is an indicator function defined as 1 if att_{xy} obeys all conditions specified in the above definition; 0 otherwise.

Example of use. Fig. 4 shows the average conformance of the metadata records in the catalogs of the NDLTD Union Archive, relative to the ETD-MS metadata standard for electronic theses and dissertations. [17] ETD-MS, different from the Dublin Core in which all fields are optional, defines six mandatory fields: dc.title, dc.creator, dc.subject, dc.date, dc.type, dc.identifier. Also the range for the dc.type is defined as the set {'Collection', 'Dataset', 'Event', 'Image', 'InteractiveResource', 'Software', 'Sound', 'Text', 'PhysicalObject', 'Stilllmage', 'Movinglmage', 'Electronic Thesis or Dissertation'}. If any value other than these words/phrases is used for the attribute, it is defined as non-conformant.

[17] http://www.ndltd.org/standards/metadata/current.html

5. Collection, metadata catalog, and repository

5.1. Collection completeness

A complete DL collection is one which contains all the existing digital objects that it should hold. Measuring completeness of a collection can be extremely hard or virtually impossible in many cases when there is no way to determine the ideal real-world collection such as in the Web or in hidden databases. Advanced judicious sampling or probing of alternative repositories whose completeness has been established manually can give crude estimates (Ipeirotis & Gravano, 2002). An example could be to approximate a measure of the completeness of a computer science collection of papers on a specific topic by sampling the ACM or IEEE-CS digital libraries, DBLP, and some other commercial publishers' on-line databases. In other cases such as for harvested or mirrored collections those estimates are easier to establish. More formally, Completeness (C_x) of a collection C_x, can be defined as the ratio between the size of C_x and the ideal real-world collection, i.e.,

$$Completeness(C_x) = \frac{|C_x|}{|\text{ideal collection}|}. \tag{6}$$

Example of use. The ACM Guide is a collection of bibliographic references and abstracts of works published by ACM and other publishers. The Guide can be considered a good approximation of an ideal computing collection for a number of reasons including the fact that it contains most of the different types of computing-related literature and for each type it can be considered fairly complete. For example, the set of theses in the Guide comes from Proquest-UMI, which receives copies of almost all dissertations defended in the USA or Canada; the number of technical reports is close to that of NCSTRL (http://www.ncstrl.org), the largest repository of CS technical reports, and it contains large numbers of records from many of the most important publishers in computer science (such as ACM, IEEE, Springer, Elsevier, etc). Table 6 shows the degree of completeness of several CS-related collections[18] when compared with the Guide.

5.2. Catalog completeness and consistency

The degree of completeness of a catalog DM_c for a collection C can be defined as

$$Completeness(DM_C) = 1 - \frac{\text{no. of } do\text{'s} \in C \text{ without a metadata specification}}{\text{size of the collection } C} \tag{7}$$

Since each object is unique by nature (e.g., each has a unique global handle) two different objects should not have the same metadata description. A catalog in which this occurs is therefore considered inconsistent. It should be noted, though, that an object can have more than one metadata specification (e.g., a Dublin Core and a MARC one).

Consistency, accordingly, is an indicator function defined as

- 0, if there is at least one set of metadata specifications assigned to more than one digital object;
- 1, otherwise.

Example of Use. In April 2004, the NDLTD Union catalog administered by OCLC tried to harvest data from the Brazilian Digital Library of Electronic Theses and Dissertations (BDTD). Because of problems in BDTD's implementation of the OAI protocol and problems with the Latin character set handling by OCLC, only 103 records were harvested from the repository. The BDTD collection contained 4446 records. Therefore the completeness of the harvested catalog for BDTD in OCLC would be completeness(BDTD in OCLC Union Catalog) =1 − (4446 − 103)/4446 = 0.023. Note that completeness significantly improved by 2006.

[18] All of which are subsets of the Guide. Size of the Guide = 735,429 (as of March, 2004).

Table 6
Completeness of several collections

Collection	Degree of completeness
ACM guide	1
DBLP	0.652
CITIDEL(DBLP(partial) + ACM(partial) + NCSTRL + NDLTDCS)	0.467
IEEE-DL	0.168
ACM-DL	0.146

5.3. Repository completeness and consistency

A repository is complete if it contains all collections it should have. The degree of completeness of a repository R is defined as

$$Completeness(R) = \frac{\text{number of collections in the repository}}{\text{ideal number of collections}}. \tag{8}$$

If the repository *stores* collections with their respective metadata catalogs, its consistency can be defined in terms of these two components. Therefore, repository consistency is an indicator function defined as

- 1, if the consistency of all catalogs with respect to their described collections is 1;
- 0, otherwise.

Example of use. We will use the ACM Guide as the ideal collection. Not considering the Bibliography and Play subcollections of the Guide and considering each publisher as a different subcollection, the completeness of CITIDEL can be calculated as 4 (ACM + IEEE + NCTRL + NDLTD-CS)/11 (total number of collections) or 0.36.

6. DL services

Dimensions of quality for DL services can be classified as external or internal (Wand & Wang, 1996). The external view is related to information satisfaction services and is concerned with the use and perceived value of these services from the point of view of societies of end users. The internal view addresses the construction and operation necessary to attain the required functionality given a set of requirements that reflect the external view. Issues in system construction, operation, design, and implementation should be considered here.

6.1. Effectiveness and efficiency

The two most obvious external quality indicators of DL services, as perceived by end users, are efficiency and effectiveness. Efficiency is most commonly measured in terms of speed, i.e., the difference between the times for request and response. More formally, let $t(e)$ be the time of an event e, and let e_{ix} and e_{fx} be the initial and the final events of scenario sc_x in service Se. The efficiency of service Se is defined as

$$Efficiency(Se) = \frac{1}{\max\limits_{sc_x \in Se}(t(e_{fx}) - t(e_{ix})) + 1.0}. \tag{9}$$

Effectiveness is normally related to information satisfaction services and can be measured by batch experiments with test collections or through experiments with real users. Different types of information services can use different metrics, the most common ones being precision and recall (Baeza-Yates & Ribeiro-Neto, 1999), extensively used to assess quality of searching or filtering services.

6.2. Extensibility and reusability

Regarding design and implementation of DL services, there are two main classes of quality properties: (1) those regarding composability of services; and (2) those regarding qualitative aspects of the models and implementations. The latter include issues such as completeness, consistency, correctness, and soundness. In this work we will concentrate on composability aspects but we acknowledge the importance and complexity of the latter issues.

Composability can be defined in terms of reusability and extensibility. In short, a service *Y reuses* a service *X* if the behavior of *Y* incorporates the behavior of *X* (in the sense that scenarios of *X* are also scenarios of *Y*). A service *Y extends* a service *X* if it subsumes the behavior of X and potentially includes additional conditional subflows of events (the scenarios of *X* are subsequences of the scenarios of *Y*). A composed service either extends or reuses another service. A composable service (i.e., a service that can be extended or reused) has to satisfy a number of requirements including exporting clear interfaces, providing mechanisms/protocols for connections and passing of parameters, offering gateway or mediator services to convert between disparate document formats and protocols, and satisfying circumstantial conditions such as satisfaction of any pre-condition based on the service's current state and input values to any called service. All of these make it very hard to quantify the composability of a service. However, even if an indicator of composability can be determined, a service is still only potentially reusable and extensible. One more pragmatic indicator of the actual composability is to ascertain from a set of services and service managers that run or implement those services, which managers are actually inherited from or included by others. Therefore given a set of services *Serv* and a set of service managers *SM* that run those services, two quality indicators of extensibility and reusability can be defined.

- $\text{Macro-Extensibility(Serv)} = \frac{\sum_{Se_i \in Serv} extended(Se_i)}{|Serv|}$, where *Serv* is the set of services of the DL and $extended(Se_i)$ is an indicator function defined as
  1, if $\exists Se_j \in Serv : Se_j$ extends Se_i;
  0, otherwise.
- $\text{Micro-Extensibility(Serv)} = \frac{\sum_{sm_x \in SM, Se_i \in Serv} LOC(sm_x) * extended(Se_i)}{\sum_{sm \in SM} LOC(sm)}$, where *LOC* corresponds to the number of lines of code of all operations of a service manager and sm_x *runs* Se_i.
- Since reuse/inclusion has a different semantics of extension, reusability can accordingly be defined as $\text{Macro-Reusability(Serv)} = \frac{\sum_{Se_i \in Serv} reused(Se_i)}{|Serv|}$, where $reused(Se_i)$ is an indicator function defined as
  1, if $\exists Se_j \in Serv : Se_j$ reuses Se_i;
  0, otherwise.
- $\text{Micro-Reusability(Serv)} = \frac{\sum_{sm_x \in SM, Se_i \in Serv} LOC(sm_x) * reused(Se_i)}{\sum_{sm \in SM} LOC(sm)}$, where *LOC* corresponds to the number of lines of code of all operations of a service manager and sm_x *runs* Se_i.

Example of use. Table 7 shows the lines of code (*LOC*) needed to implement service managers that run several services in the ETANA archaeological digital library, in September 2004 (Ravindranathan, 2004). Let us assume a 1:1 ratio between the set of services and the set of service managers. Reused services (and included service managers) are implemented as ODL components (Suleman, 2003). These services are searching, annotating, recommending, and (union) cataloging.

The wrapping services, the ones that really reuse and provide the services offered by the DL components, are necessary in order to deal with issues such as invoking operations, parsing results, and interfacing with other components (like the user interface). However, the additional code for those wrappers is only a very small percentage of the total lines of code required for implementing the components. In the ETANA-DL prototype (in September 2004), only a few important services were componentized and therefore reused Macro-Reusability(ETANA DL Services) = 4/16 = 0.25. However, Micro-Reusability = 3630/11,910 = 0.304 makes it clear that we can re-use a very significant percentage of DL code by implementing common DL services as components. Moreover, as more service managers are componentized, more code and managers are potentially inherited from/included by more DLs.

6.3. Reliability

Regarding operation, the most important quality criterion is reliability. Service reliability can be defined as the probability that the service will not fail during a given period of time (Hansen, 1983). We define the reliability of a service Se_x as

Table 7
Analysis of ETANA DL prototype using the metric of lines of code

Service	Component based	LOC for implementing service	Total LOC	LOC reused from component
Searching – back-end	Yes	–	1650	1650
Search wrapping	No	100	100	–
Recommending	Yes	–	700	700
Recommend wrapping	No	200	200	–
Annotating – back-end	Yes	50	600	600
Annotate wrapping	No	50	50	–
Union catalog	Yes	–	680	680
User interface service	No	1800	1600	–
Browsing	No	1390	1390	–
Comparing (objects)	No	650	650	–
Marking Items	No	550	550	–
Items of interest	No	480	480	–
Recent searches/discussions	No	230	230	–
Collections description	No	250	250	–
User management	No	600	600	–
Framework code	No	2000	2000	–
	Total	8280	11,910	3630

Table 8
Reliability of CITIDEL services

CITIDEL service	No. of failures/No. of accesses	Reliability
Searching	73/14,370	0.994
Browsing	4130/153,369	0.973
Requesting (getobject)	1569/318,036	0.995
Structured searching	214/752	0.66
Contributing	0/980	1

$$Reliability(Se_x) = 1 - \frac{\text{no. of failure}}{\text{no. of accesses}}. \tag{10}$$

A failure is characterized as an event that

(1) was supposed to happen in a scenario but did not, or
(2) did happen but did not execute some of its operations, or
(3) did happen, where the operations were executed, but the results were not the correct ones.

Example of use. Table 8 shows reliability figures for the most popular services of CITIDEL, according to a log analysis done on April 1, 2004. The low reliability for the *structured searching* service can be explained by the fact that it was experimental, and ran only for a short period of time. However, entry points and links to this service were not removed after the experiments, and users kept using it without getting answers. This also shows how flaws in design can be found with such quality-oriented analysis.

7. Quality and the information life cycle

Given the fact that information in digital libraries is carried by digital objects and their respective metadata specifications, the proposed dimensions of quality for these two concepts can be connected to the life cycle of information in digital libraries (Borgman, 1996).[19] Such connections can be used to determine when and where

[19] Again, since most of the key information content in a DL is carried by these 2 concepts (digital objects and metadata specifications) the focus here is on them rather than all the other parts of a DL. An expansion to cover others aspects is part of future work.

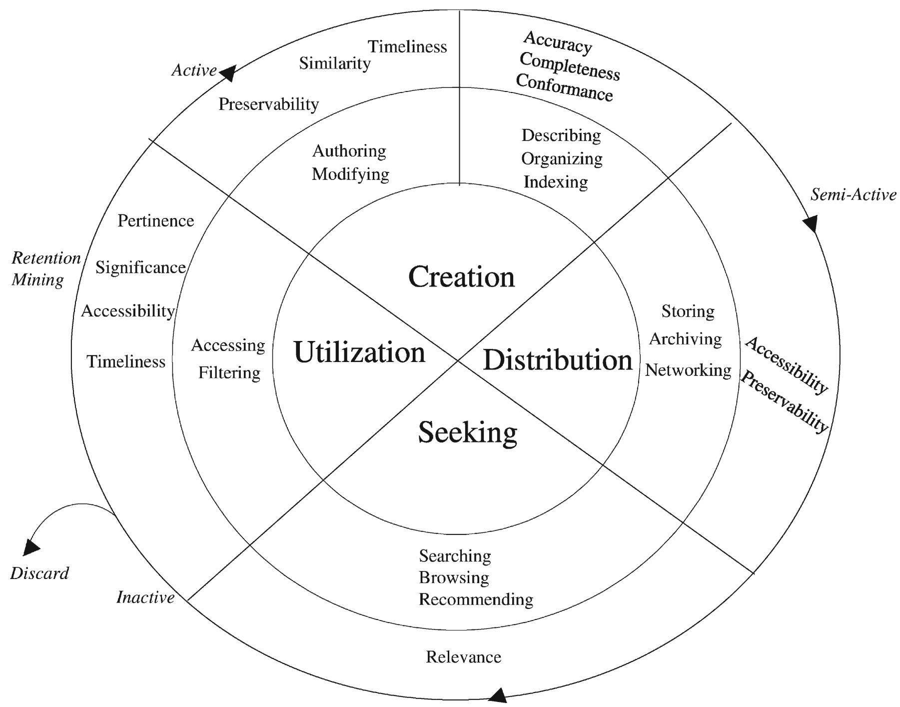


Fig. 5. Information life cycle (adapted from Borgman (1996)) with respective dimensions of quality added for each major phase and related activities.

quality indicators can be measured, assessed, and improved – as well as how possible quality problems can be prevented, detected, and eliminated.

The connections are shown in Fig. 5. The life cycle (see inner portion) has four major phases: information creation, distribution, seeking, and utilization. The outer arrows show in which stage information is active, semi-active, or inactive with regard to the phases. Each major phase (see inner ring) is connected to a number of activities or services. Finally (see outer ring), each dimension of quality is associated with a corresponding set of activities/services.

Similarity to other information, digital objects, or versions can be assessed at time of creation and modification. Preservability and timeliness (in relation to modifications) also are related to this phase. The next subphase in the cycle deals with metadata creation and information organization and description; therefore all quality dimensions associated with metadata specifications are located here. Special metadata quality processes such as enforcing filling out of mandatory fields and use of specific formats (e.g., for dates) as well schema validations, should be applied to related activities to guarantee quality (namely, accuracy, completeness, and conformance).

In the next phase of archiving and distribution, the aspects of accessibility and preservability (e.g., means of storage, format, position in an organizational scheme, etc.) should be taken into consideration. In the seeking (e.g., searching) phase, relevance of information as returned by the several information satisfaction services can be measured. Finally, most of the dimensions associated with the perceived value of the information (pertinence, significance) can be assessed during the utilization phase.

8. Evaluation: focus group

In order to assess the potential practical utility of the proposed quality model in the library and information science (LIS) world, we conducted a focus group with three librarians with experience in practical library work and digital libraries.

This focus group meeting included a presentation of approximately 60 min duration about 5S and the proposed quality indicators (with examples) by the researchers/moderators, followed by a 30 min discussion.

Questions, comments, and discussions were allowed during the presentation. In particular, the discussions were focused around four questions:

- Are you able to understand the 5S model?
- How does it relate to (your) library world?
- How do the proposed indicators relate to your practices in the library?
- Would you be willing to apply these measures to your (digital) libraries?

8.1. Presentation – discussions

Discussions during the presentation centered around 5S itself and the utility of the quality model. The discussion started with the basic question "What's a digital library?" that was raised by one of the participants. The moderator emphasized that to precisely define digital libraries was one of the main goals of the work.

When first confronted with 5S, some of the participants felt that the framework made sense, but the concepts of scenarios and spaces were the least intuitive. Regarding streams, it was felt that intuitively they seem active and dynamic, but the ones used in this work were more static, which was a bit counter intuitive. One of the participants also said that she had a hard time seeing how the 5S terminology maps to the concepts normally used in the library world. A mapping of conventional library and information science terms into the 5S framework was suggested. For example, another participant raised the question "what does a database mean in this framework?". After learning that for that participant a database means a storage of documents, we concluded that it corresponded to the 5S notions of collections and repositories.

One of the most discussed aspects was the "minimalist" approach we took in this work. There was almost a consensus that this was the right way to go. One of the participants suggested adding "reference service" – essential in the library world – to the minimal DL. This raised concerns that we were dealing with DLs as something outside of the library: "shouldn't the DL be part of the library?" asked one of the participants. It was explained why in a field with so many constructs we could be more elegant and precise to clarify terms with our approach. In the end it was the consensus that the framework is OK, if the focus is on DLs, not (conventional) libraries. Another participant also raised concerns about the higher level constructs being understandable by the general public and librarians. Finally some of the relationships among the services (e.g., between indexing and searching) and additional (traditional) DL services were discussed.

In the second phase of the presentation, discussion centered around the quality model. One of the first questions raised was the difference between pertinence and relevance. The explanation was accepted by the participants. Another question which drove part of the discussion was "what do you mean by 'good'?" and "by good you mean only from the user perspective?" This shifted the presentation to the connections between the proposed quality indicators and the several phases of the information life cycle.

The discussion then shifted to each proposed quality indicator. Participants seemed to agree with most of the dimensions and definitions. First concerns were raised regarding the "timeliness" indicator for digital objects. One participant argued that in some instances it is impossible to find the age of every object. It was proposed to have a categorization for "ageless" objects – those never reviewed, reviewed in 5 years, etc. The issue of the age of digital objects that are surrogates for real world artifacts also was discussed. In the context of collection-related indicators, a participant suggested that some of these quality measures could be used for collection management mainly as selection criteria to "get rid of things" (as stated in her own words). There was a consensus that practical aspects of dealing with storage space and limited resources are rarely discussed in the field.

The strongest reactions were generated by the issues of catalog and collection completeness. It was thought that in some cases, for instance catalogs based on the Dublin Core (15 attributes), this indicator made sense, but in the case of MARC (with hundreds of attributes) it would not. It was felt that the context in which those indicators would be more applicable should be elaborated upon and that these indicators in some cases had more theoretical value than practical application. In the particular case study presented (viz., OCLC NDLTD Union Catalog), it was conjectured whether the low values for completeness of one catalog – MIT – were due to the fact that most objects in the respective collection were scanned, not born digital. Finally, since we had many indicators for each concept, one argued that we really did need several indicators to get something meaningful (e.g., practically useful indicators might be computable based on some function applied to some combination of several of our indicators).

8.2. Post-presentation discussion

The post-presentation discussion started with one of the moderators encouraging the participants to think out loud about how they felt about the whole thing. One of the participants started by expressing his view that the two pieces of the presentation were a bit disconnected, i.e., he felt that the quality model was not very much associated with 5S. He suggested that if the discussions were viewed like a novel, he would think that the "5S" character had disappeared when quality was considered.[20] The same participant expressed his view that we were proposing two kinds of indicators: (1) some measured against a perfect thing that cannot always be known or defined; and (2) others that are highly specific, arguably useful. In the real world, something in between would be the ideal. The criticism was not restricted to our study, but also applied to most similar studies and other types of quality measures he knows and applies in the library in his work. It was suggested that expressing the contexts where these indicators would be more or less useful would help in their adoption.

Another important question was raised by another participant. In her own words: "What do I do with these measures?". It was agreed that the goal is to promote improvements and make things better. It also was suggested that these could be used somehow in training of digital librarians and DL administrators. Thus, we use these in a curriculum development project launched in 2006 (Pomerantz, Oh, Wildemuth, Yang, & Fox, 2006).

One of the moderators shifted the discussion to the broader questions of "Which ideas apply to LIS? Which to DL? Which to both?" The first reaction was that DL is only a part of the library; some things cannot be digitized. Another participant said that it helped to hear about our minimalist focus – it would help acceptance within LIS. The same participant expressed the personal opinion that 5S would not be much use to him but it might be very helpful to IT people involved with (digital) library issues; impact on LIS was uncertain. Another suggested a study to correlate user satisfaction with the quality indicators. For the explanation of the indicators themselves it was suggested that the application of all of them in a running example would be really helpful. Regarding a proposed quality toolkit, one participant felt more work on services, including infra-structure services and preservation, was necessary.[21] The suggestion was that we needed to go beyond a minimal DL for this work to be practically useful.

In the end, everybody felt the work was very interesting and useful, with potential to help the field of *digital* libraries. In the words of one of the participants who was a little bit familiar with the 5S framework: "you have come a long way with 5S and that is extremely impressive, but more needs to be done before this thing gets widely accepted and practically used".

9. Related work

DL quality and evaluation is a very underrepresented research area in the digital library literature. Saracevic (2000) was one of the first to consider the problem. He argued that any evaluation has to consider a number of issues such as the context of evaluation, the criteria, the measures/indicators, and the methodology. However, his analysis concluded that there are no clear agreements regarding the elements of criteria, measures/indicators, and methodologies for DL evaluation. In an attempt to fill some gaps in this area, Fuhr et al. (2001) proposed a descriptive scheme for DLs based on four dimensions: data/collection, system/technology, users, and usage. We see the work proposed in this paper regarding DL quality issues and evaluation as a next, complementary step in this area, one that is based on a sound, formal theory for DLs.

Papers in a workshop on "Evaluation of Digital Libraries" touched on some of the issues discussed here. Quoting one of the papers in the proceedings of that workshop (Agosti, Nunzio, & Ferro, 2004): "thus it could be worth discussing whether the 5S is an appropriate model for facing this kind of (evaluation) issue and

[20] In the interest of time, the connection between 5S theory and quality, as explained in this paper, could only the touched on during the presentation.

[21] Because of time constraints, the taxonomy of DL services was not presented to this group. Otherwise, this objection might not have arisen.

whether it could further a better understanding in this research field''. That is exactly what we have tried to do in this paper. We are encouraged to continue this work, and collaborate with others on this regard, as exemplified by recent activities coordinated as part of DELOS efforts.

10. Conclusion

In this article, we have proposed a quality model for digital libraries that is comprised of a number of quality indicators for key formal DL concepts. The utilization of many of these indicators was illustrated with real case studies. Connections with the Information Life Cycle and a focus-group-based evaluation of the model were also discussed.

This work may help current managers of digital libraries to begin to develop a more quantitative assessment program. It may help designers of new digital libraries to identify constraints and tradeoffs needed to ensure adequate quality, to plan how to monitor system behavior in order to facilitate evaluation, and to set priorities that will help ensure desired levels of performance as well as desired collection characteristics. It may help developers of new digital library systems, as they build in means to measure and ensure system quality, as well as operational quality of service.

In future work we hope to extend our efforts toward developing a quality toolkit for digital libraries that can be deployed for pilot quality assessment studies. We hope to collaborate with others in extending the set of quality indicators, in providing more formula, in refining the current set of mostly system-related dimensions, in adding in usage-related dimensions, and in connecting this work with logging and log analysis activities. On the usage-related side, we will consider popularity of scenarios and digital objects, correctness of scenario models, usability of services, and educational potential of resources (Sumner, Khoo, Recker, & Marlino, 2003). Finally, we hope to connect this research with the development of curriculum materials that will be helpful for teaching and learning about digital libraries (Pomerantz et al., 2006).

Acknowledgements

Thanks go to sponsors of our work, which include: AFOSR (Grant F49620-02-1-0090), AOL, IMLS, NSF (Grants DUE-0121679, IIS-0325579, IIS-0535057) and CNPq project 5S-VQ (Grant MCT/CNPq/CT-INFO 551013/2005-2). Thanks also go to those in Virginia Tech's Digital Library Research Laboratory, and to all those involved in our various related projects.

References

Agosti, M., Nunzio, G. M. D., & Ferro, N. (2004). Evaluation of a digital library system. *Workshop on the evaluation of digital libraries, DELOS WP7* (pp. 73–76). Padova, Italy.

Amsler, R. (1972). Application of citation-based automatic classification. *Technical report*. The University of Texas at Austin, Linguistics Research Center.

Babbie, E. (1990). *The practice of social research* (6th ed.). Belmont, CA: Wadsworth.

Baeza-Yates, R., & Ribeiro-Neto, B. (1999). *Modern information retrieval*. New York: Addison-Wesley.

Borgman, C. L. (1996). Social aspects of digital libraries. In E.A. Fox, & G. Marchionini (Eds.), *Proceedings of the 1st ACM international conference on digital libraries* (pp. 170–171). Bethesda, MD.

Cosijn, E., & Ingwersen, P. (2000). Dimensions of relevance. *Information Processing and Management, 36*(4), 533–550.

Crespo, A., & Garcia-Molina, H. (1998). Archival storage for digital libraries. In I. Witten, R. Akscyn, & M. F. Shipman (Eds.), *Proceedings of the 3rd ACM international conference on digital libraries* (pp. 69–78). Pittsburgh, PA.

Foskett, D. J. (1972). A note on the concept of relevance. *Information Storage and Retrieval, 8*(2), 77–78.

Fuhr, N., Hansen, P., Mabe, M., Micsik, A., & Sölvberg, I. (2001). Digital libraries: a generic classification and evaluation scheme. In P. Constantopoulos, & I. Sölvberg (Eds.), *Proceedings of ECDL'01* (pp. 187–199). Darmstadt, Germany.

Gonçalves, M. A. (2004). Streams, structures, spaces, scenarios, and societies: a formal framework for digital libraries and its applications: Ph.D. thesis, Virginia Tech CS Department. http://scholar.lib.vt.edu/theses/available/etd-12052004-135923.

Gonçalves, M. A., Fox, E. A., Watson, L. T., & Kipp, N. (2004). Streams, structures, spaces, scenarios, societies (5S): a formal model for digital libraries. *ACM Transactions on Information Systems, 22*(2), 270–312.

Greisdorf, H. (2003). Relevance thresholds: a multi-stage predictive model of how users evaluate information. *Information Processing and Management, 39*(3), 403–423.

Hansen, J. V. (1983). Audit considerations in distributed processing systems. *Communications of the ACM, 26*(8), 562–569.

Hunter, J., Choudhury, S. (2003). Implementing preservation strategies for complex multimedia objects. In T. Koch, & I. Sölvberg (Eds.), *Proceedings of the 7th European conference on research and advanced technology for digital libraries, ECDL 2003* (pp. 473–486). Trodheim, Norway.

Hunter, J., Choudhury, S. (2004). A semi-automated digital preservation system based on semantic web services. In H. Chen, H. D. Wactlar, C. Chen, E. Lim, & M. G. Christel (Eds.), *Proceedings of the fourth ACM/IEEE-CS joint conference on digital libraries* (pp. 269–278). Tucson, Arizona.

Ingwersen, P., van Rijsbergen, K., & Belkin, N. (2004). *ACM SIGIR 2004 workshop on "Information Retrieval in Context"*. Glasgow Information Retrieval Group. http://ir.dcs.gla.ac.uk/context/IRinContext_WorkshopNotes_SIGIR2004.

Ipeirotis, P. G., & Gravano, L. (2002). Distributed search over the hidden Web: hierarchical database sampling and selection. In S. Bressan, A. B. Chaudhri, M. Lee, J. X. Yu, & ZoLacroix (Eds.), *Proceedings of the 28th international conference on very large data bases* (pp. 394–405). Hong Kong SAR, China.

Kemp, D. A. (1974). Relevance, pertinence, and information system development. *Information Storage and Retrieval, 10*(2), 37–47.

Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American Documentation, 14*(1), 10–25.

Levy, D. M. (1998). Heroic measures: reflections on the possibility and purpose of digital preservation. In I. Witten, R. Akscyn & K. M., Shipman (Eds.), *Proceedings of the 3rd ACM international conference on digital libraries* (pp. 152–161). Pittsburgh, PA.

Lorie, R. A. (2001). Long term preservation of digital information. In E. A. Fox, & C. L. Borgman (Eds.), *Proceedings of the 1st joint conference on digital libraries* (pp. 346–352). Roanoke, VA.

Mizzaro, S. (1996). A cognitive analysis of information retrieval. In P. Ingwersen, & N. O. Pors (Eds), *Information science: integration in perspective – proceedings of CoLIS2* (pp. 233–250). Copenhagen, Denmark.

Mizzaro, S. (1998). How many relevances in information retrieval? *Interacting With Computers, 10*(3), 305–322.

Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communications of the ACM, 45*(4), 211–218.

Pomerantz, J., Oh, S., Wildemuth, B. M., Yang, S., & Fox, E. A. (2006). Core: digital library education in library and information science programs. *DLib Magazine, 12*(11).

Ravindranathan, U. (2004). Prototyping digital libraries handling heterogeneous data sources – an ETANA-DL case study. *Master's thesis*, Virginia Tech CS Department.

Redman, T. C. (1992). *Data quality – management and technology*. New York: Bantam Book.

Rothenberg, J. (2000). *Using emulation to preserve digital documents*. The Netherlands: Koninklijke Bibliotheek.

Rowan, M., Gregor, P., Sloan, D., & Booth, P. (2000). Evaluating web resources for disability access. In M. Tremaine, E. Cole, & E. Mynatt (Eds.), *Fourth annual ACM conference on assistive technologies* (pp. 80–84). Arlington, VA.

Sanett, S. (2003). The cost to preserve authentic electronic records in perpetuity: comparing costs across cost models and cost frameworks. *RLG DigiNews, 7*(4)http://www.rlg.org/preserv/diginews/v7_n4_feature2.html.

Saracevic, T. (1975). Relevance: a review and a framework for thinking on the notion in information science. *Journal of the American Society for Information Science, 26*(6), 321–343.

Saracevic, T. (1996). Relevance reconsidered. *Proceedings of the Second International Conference on Conceptions of Library and Information Science, 39*(3), 201–218.

Saracevic, T. (2000). Digital library evaluation: toward evolution of concepts. *Library Trends, 49*(2), 350–369.

Sayood, K. (1996). *Introduction to data compression*. San Mateo, CA: Morgan Kaufman Publishers.

Small, H. G. (1973). Co-citation in the scientific literature: a new measure of the relationship between two documents. *Journal of the American Society for Information Science, 24*(4), 265–269.

Suleman, H. (2003). Open digital libraries. *Ph.D. thesis*, Virginia Tech Department of Computer Science, http://scholar.lib.vt.edu/theses/available/etd-11222002-155624/.

Sumner, T., Khoo, M., Recker, M., & Marlino, M. (2003). Understanding educator perceptions of "quality" in digital libraries. In L. Delcambre, G. Henry, & C. C. Marshall (Eds.), *Proceedings of the 3rd ACM/IEEE-CS joint conference on digital libraries* (pp. 269–279). Houston, TX.

Taylor, R. S. (1968). Question-negotiation and information seeking in libraries. *College and Research Libraries, 29*, 178–194.

Voorhees, E. M. (2001). Evaluation by highly relevant documents. In D. H. Kraft, W. B. Croft, D. J. Harper, & J. Zobel (Eds.), *Proceedings of the 24th annual international ACM SIGIR conference on research and development in information retrieval* (pp. 74–82).

Wand, Y., & Wang, R. Y. (1996). Anchoring data quality dimensions in ontological foundations. *Communications of the ACM, 39*(11), 86–95.

Waugh, A., Wilkinson, R., Hills, B., & Dell'oro, J. (2000). Preserving digital information forever. In P. J. Nurnberg, D. L. Hicks, & R. Furuta (Eds.), *Proceedings of the 5th ACM international conference on digital libraries* (pp. 175–184). San Antonio, TX.

Zhang, B., Gonçalves, M. A., & Fox, E. A. (2003). An OAI-based filtering service for CITIDEL from NDLTD. In T. M. T. Sembok, H. B. Zaman, H. Chen, S. Urs, & S. H. Myaeng (Eds.), *Proceedings of the 6th international conference on Asian digital libraries, ICADL 2003* (pp. 590–601). Kuala Lumpur, Malaysia.