

# Quality Criteria for Digital Information in Long-term Digital Preservation

Margarita Korenkova and Ann Hägerfors, Luleå University of Technology; Luleå, Sweden

## Abstract

*Preservation of digital information is a vexing problem for preservation agencies. Properties or functionalities of the preserved object can be lost over time. Digital preservation is an ongoing, proactive process of preserving information and its significance over time. In this article, we conduct a theoretical and an empirical study in Swedish preservation agencies to identify and define a set of essential information properties for long-term preservation. Among identified properties, the study confirmed that knowledge on how to guarantee quality of preserved digital information is low and showed that guiding quality principles need to be established. The need for a unified information quality assurance framework was also highly confirmed.*

## Introduction

Digital preservation is quickly growing into a critical and vexing problem for preservation agencies. The root of the problem is the quickly evolving technology as well as the ever-growing amount of digital information to preserve. Preserving something digitally does mean that some properties or functionalities of the preserved object are eventually lost, either through conversion or through changing technology [3] [5] [7]. Even if we could assure the preservation of digital resources and overcome media fragility and technological obsolescence, preserved materials will be of little or no value unless quality can be guaranteed over time. During preservation the quality of information properties need to be controlled in order to assure the quality of information. For this to be possible, a set of essential information properties, “significant properties”, for continuous information quality assurance need to be defined.

Archives and libraries lack sustainable solutions, strategies, methods, knowledge and experience to handle preservation of digital material and there is a need for a common foundation concerning terms and concepts required for development of sustainable and financially sound high quality solutions [6] [27] [29] [11] [22].

In this article, we identify a set of properties for information quality that are proposed as essential for long-term control and maintenance based on a literature study on quality assurance and an empirical study on information quality requirements. The identification of significant information properties based on the quality perspective is deemed important for further research on measurement and maintenance of significant properties quality, eventually leading to a quality framework for digital preservation. We searched for quality criteria or significant properties based on an information perspective.

## Method

This article is based on a literature study and an interview study with archives and library professionals. For the literature study, literature was chosen based on appropriateness to long-term

digital preservation. It covers Total Quality Management, Archival science, Library and Information Science; Information Systems Science, Digital Preservation; Records Management and Quality Improvement.

Some quality criteria were used as synonyms or with overlapping characteristics from other criteria. In the analysis we have focused clarification of the meaning of the properties in order to display fairly clearly differences between them.

The interview study was carried out at the Swedish National Archives and the National Library of Sweden, which are the foremost active in Sweden concerning digital preservation. We interviewed a total of nine respondents, working with digital preservation, five persons in the National Archives and four persons in the National Library. The interviews were semi-structured [34].

## Quality Criteria

The research was carried out within the Long-term Digital Preservation Centre at Luleå University of Technology (the LDP-Centre), where digital preservation is viewed as an ongoing, proactive process of preserving information and its significance over time through maintaining information properties. The primary goal of long-term digital preservation is in this perspective to preserve information for future users. It is information that has meaning and is meaningful to human users. Users are interested in informational content and its meaning (the intellectual property of digital information), rather than in the technical side of it even if technological solutions are necessary in order to make digital information available and accessible. The information perspective focuses on the future user and the survival of preserved information over changing technologies. Information is defined as data and descriptive information that allows the material to be understood by humans over time. [22] [23] [21] [28] [27]

In the literature study, we found a stated need for quality models in digital preservation as a corner stone of research or development as well as a number of attempts to address the issue. Williamson [33] called for awareness and implementation of appropriate quality assurance procedures at each stage in the process of digital curation in order to maximize the return of investment being made in digital curation.

Within Information Systems Science, the research has been focused on studies of success of information systems (IS), e.g., measurement of how certain requirements or qualities have been reached by a certain information system [1]. For instance, Holmes [10] has examined ten attributes of quality of data in IS as benchmarks to improve the effectiveness of IS in the business organizations: accuracy, timeliness, completeness, coherence, format, accessibility, compatibility, security, validity. Liew & Foo [19] proposed a set of properties in the information objects within the interaction environment to support the enhanced interaction and value-adding of electronic documents, such as electronic

journals: structured, contextualized, explicable, quireable and navigable and affiliated with layers of additional information and metadata.

The concept of data quality dimensions has been explored by Richard Wang and the Total Data Quality Management group at the Massachusetts Institute of Technology from an enterprise perspective and organizational data quality improvement. [30] [31] [16] [24] [18] Wang and Strong [31] proposed 20 data dimensions grouped in four target categories. *Accuracy of data* consists of accuracy, objectivity, believability, completeness, traceability, reputation and variety of data sources. *Relevancy of data* includes value-added, relevancy, timeliness, ease of operation, flexibility and appropriate amount of data. *Representation of data* consists of interpretability, ease of understanding, representational consistency and concise representation. *Accessibility to data* includes accessibility, cost-effectiveness and access security.

Data quality in large data sets or databases is comprehensively covered in [8] [20] [25]. Data quality is explained in several dimensions that make data (existing in data bases) appropriate for a specific use.

Borglund [2] studied quality in recordkeeping and information systems design and proposed a Recordkeeping Quality Assessment Model (RQAM), which aims to support assessment and measurement of quality of records in recordkeeping systems. The study states that electronic records have unique quality dimensions, but quality dimensions used for information and data are also usable for electronic records.

Within Long-term Digital Preservation, so far, much research has been limited to technical aspects of preservation [21] [22] [2] [28]. Dollar [5] proposed eight criteria for long-term digital preservation of records in archival context: readable, intelligible, identifiable, encapsulated, retrievable, reconstructable, understandable and authentic. The Swedish law states that digital archival records to be preserved are so in close to its original form or unchanged and prepared to be reconstructed in its original context, i.e., it implies preservation of the ability to reproduce the record [28]. Nilsson [21] proposed essential criteria based on Dollar [5] in long-term digital preservation field: trustworthiness, usefulness and understandability, and accessibility and availability. Another outcome of Nilsson's work was a "Mental Model" of key concept for digital information object in long-term digital preservation. The Council on Library and Information Recourses [3] and Duranti [7] have addressed user's criteria: accuracy which implies that information is precise, free-of-error or distortions and authentic. Authenticity which implies that the identity and integrity of information resources have not been inadvertently or maliciously compromised, and they are what they purport to be. According to Williamson [33], criteria of accessibility and integrity of digital information are central to the process of digital preservation. Kelton et al., [17] proposed a "Model of Trust" and identified: trustworthiness, confidence, reliability, accuracy, currency, coverage, believability, credibility, objectivity, validity and predictability.

Recent work on quality research proposes research in information quality assessment, information quality management and contextual information quality [9].

The Quality Working Group of the DL.org proposes a model that embraces quality parameter such as Generic Quality, Content Quality, and Policy Quality. The aim is to investigate quality

measurements regarding Digital Libraries as "Organizations" including Digital Library System and Digital Library Management. [4] This is the nearest work to digital preservation.

The goal of the InSPECT project was to propose a framework for guiding the process of identifying, analyzing and recording the elements and essential technical properties of an Information Preservation Object that are necessary for manifestation and recreation of a digital object identify that are necessary to recreate these information objects in the future. This framework is useful to analyze the object and obtain a complete list of technical properties associated with the Information Object. The project used five high-level categories distinguishing properties of the Information Object from those of the Data Object: content; context, rendering, structure and behavior [13]

According to Joint Information Systems Committee [15], significant properties are referred to as "significant characteristics" or "essence". They are essential attributes of digital object which affect its appearance, behavior, quality and usability. Significant properties must be preserved over time for the digital information to remain accessible and meaningful. Significant properties are identified to be requirements for quality of digital object, which can be preserved anywhere, e.g. on some storage media, or in an information system, rather than criteria for information systems.

It is reasonable to assume that some properties are common and can be applied to several types of information. The level of significance of other properties is likely to be dependent on different contexts of creation and use. In some cases, for example, it may be decided that the textual content of a document is the most important element. If the "look and feel" of the original document forms part of its intellectual content, then the "look and feel" needs to be preserved [21]. Proper understanding of the significant properties of digital information as well as identification of both common and specific properties is critical to establish best practice approaches to digital preservation. It assists appraisal and selection processes in which choices are made about which significant properties of digital information is worth preserving; it helps the development of preservation metadata, the assessment of different preservation strategies and informs future work on developing common standards across the preservation community.

To summarize, we found the following criteria that are essential from the information perspective. Technological criteria are prerequisites for these to be implemented, but are not mentioned here since the aim is to identify technology independent criteria. As we have stated before, in our perspective information is at the core of preservation. Information is a living concept and, if treated properly, has the potential to live longer than people, technological means and organizations. We have identified those criteria that are in direct relation to humans understanding of the preserved information. This means that for example the concepts "accessible" and "available" are not included since they refer to humans and their interaction with technology.

**Accuracy:** Information is reliable, and certified free of error. **Reliability** means unaltered, unchanged and uncorrupted. This implies full and accurate representation of the transactions, activities or facts. Reliability incorporates confidence, which plays a role in users attitudes toward digital information. **Free-of-error** means that information is correct, unchanged and reliable over time.

**Affiliation with metadata and additional layer of information:** Provides additional information about the content, for instance evaluation, reviews, commentaries and criticism.

**Authenticity:** Persistence of its original or faithful to original characteristics over time. Content, context and structure of the record remains in a reliable state.

**Coherence:** The information forms a consistent whole.

**Completeness:** Information is of sufficient breadth, depth, and scope for the task at hand and no piece of information is missing.

**Conciseness:** Information is compactly represented without being overwhelming i.e., brief in presentation, yet complete and to the point.

**Contextualized:** The property that demonstrates the relationships between information resources and the environment in which they were created, are now managed and preserved as long as needed.

**Fractional:** Users are able to view, navigate, search and manipulate data through the structured object at the all levels, e.g., collections, documents, paragraphs, sentences or words.

**Identifiable:** In order to be understandable for humans the information object needs to be possible to identify and distinguish from other information objects in the system.

**Interpretable:** Data are in appropriate language and units and the data definitions are clear. Interpretability contributes to understandability of information.

**Objectivity:** Data are unbiased (unprejudiced) and impartial.

**Readable:** Information is readable for users.

**Relevant:** Information is applicable and helpful for the task at hand, e.g. for resolving users' problems.

**Reputational:** Information is trusted or highly regarded in terms of their source or content.

**Structured:** Data object are structured on several levels, e.g., sets, collections, ordered lists, trees, tables, natural language sentences, sections, paragraphs or words.

**Timeliness:** The age of the data is appropriate for the task at hand.

**Traceability:** Information is well documented, verifiable, and easily attributed to a source. All information objects should have representations of their history – where they originate from and what operations have been performed on them. The historical sequence of the objects should be presented in a visualisable, malleable history mechanism.

**Trustworthiness:** Information in preservation object has to be trustworthy. Information has to be captured and delivered in a trusted way. The users must believe that information is reliable and trust that they got all the information they should get. The information must in such a state that it can be accepted or regarded as true, real, and credible.

**Understandability:** Information is clear without ambiguity and easily comprehended. The meaning of information is determined by its syntax, semantics, words, context of creation and use.

**Usability:** Information is available and possible to use for those who wish to do so.

**Usefulness:** Information is beneficial and provides advantage from its use.

**Validity:** Refer to aspects of ethics in use of responsible and accepted practices. This includes the soundness of methods used

during the whole lifecycle of information from creation and through the whole preservation process, e.g. verification of information, appropriate citation of sources. Information which can be verified as true and satisfying appropriate standards or principles related to other quality criteria have validity.

It is worth noting that the criteria were established for different purposes and contexts; they refer to different levels of abstraction of a digital preservation object, and have different ontological bases. The literature study showed little explicit focus on information or technology independent solutions, which in the perspective of the LDP-centre is essential for sustainable solutions.

## Empirical Study

In the interviews, the respondents talked about what quality in digital preservation means for them. In this section we focus on what they had in common, since it is common criteria we wish to identify. Specific criteria are however briefly mentioned at the end of the section.

Descriptive information is expressed in metadata. It provides meaning to captured and preserved content information. Descriptive information is divided into several groups: contextualized, preservation, administrative, structural, juridical and technical. Descriptive information is important to humans for interpretation, understanding and assimilation of information, to assure that information is authentic, intact, not altered or manipulated. The descriptive information makes it easier for curators to make digital material available and usable.

Data or information might have to be added by curators after the material has been delivered to the preservation agencies. It can be descriptions about uncertainty, unreliability or something unmatched. Descriptive information can be extended and developed by curators when new and more reliable information emerges, to make information more complete or to create documentation of managerial or administrative actions performed on the information during preservation.

Descriptive information needs to be sufficient enough to be able to interpret and understand information content by users and to be able to understand it from different angles for different needs. If descriptive information is not complete, information content will be un-interpretable, incomprehensible, meaningless, and unusable.

It is important that information should continue to be in a condition that remains as close as possible to its initial condition.

Usable information is information that is obtained from content and appearance and that is described, complete, readable, interpretable, understandable, correct, unaltered and accessible. Usability of information is determined from which potential risks are and what changes are permitted or accepted. Information that is usable, demanded and used by current and future users is useful.

Available information is data or information that exists, is accessible and permitted to use and usable. Available information needs to be prepared for search, picking out and derivation of data from a computer system for users. Preservation agencies act as a service and provide service for information users. The provided information needs to be understandable, in good condition and satisfy users. Therefore data needs to be well-documented, be in an appropriate storage format and usable.

Information needs to be viewable with the same content and components as in the initial state. Generally, logical appearance of

data is important for grouping, interpreting, understanding and making it available for users. However, physical appearance, e.g. layout or format may be a different matter. Visual appearance is the extent to which some changes in information appearance are permitted or accepted. Information content is viewed as most important to preserve, while physical appearance is difficult to preserve, as information loses its searchability quality.

Information need to be accessible. This means that users may access the information virtually from different locations. Privacy classified information need to be accessible inside the preservation agencies by assigned authorized persons. Accessible information must be usable and access needs to be permitted.

Both data and metadata need to be searchable in order to be picked up from a preservation package, put together, provided and presented to users in a readable form. For this to be assured, data need to be well-described, available and accessible.

Information need to be traceable, i.e. there is a need for traceable chains of events (e.g., conversions), continuously updated and saved. This is important in order to assess authenticity, reliability and trustworthiness of information. People should be able to browse and see the history in order to ensure that the information is not corrupted and that nothing has happened to it over time.

Information needs to be readable. Users should be able to read and understand the information, even after many format conversions. To ensure document readability, the data must be stored in a format that is simple and application-independent, the conversions must be minimized, standards need to be used and requirements need to be set today. Technical requirements need to be assessed, i.e., what is required of a platform to be able to read a document. In order to ensure readability for users, information should be showed in its initial state.

Users need to be able to rely on that information and data are correct, otherwise it is not useful. Deficiencies in reliability might be caused by lacking or incomplete metadata, errors in the field descriptions, but also by errors in data files.

The information (both data and metadata) must fulfill the criteria of not having been changed, mutilated, tampered with or sabotaged in any way.

Information needs to be presented in an easy way to be understood. Contextual information is required to understand the content. It is difficult to understand a picture or a text without knowing the context, in which it was created because it is not about an isolated snippet, but rather who use it, how to look at it, for whom is it, how to use it and which links are there.

Users have to be able to trust that information is accessible, accurate, trustworthy, complete, reliable, not forged or modified, not distorted, destroyed or tampered with in any way. Users should be able to browse and see from the history that nothing has distorted with the information, thereby verifying that the procedures and processes that are used are accurate. This includes trust in that it is the same information content, that all database records and fields contain the same information, the same data. It also entails trust in the preservation agencies.

Data and metadata should be interpretable. Particular descriptive information (metadata) is necessary to interpret the data in the future. Requirements for documentation are about the level of information, i.e., how to interpret the data so that it becomes information. For interpretability, data should be

organized in such a way and have associated metadata that allows users to interpret and understand data in its context.

Document should be stored technically, rendered, reproduced, and represented in a way (in a coherent process) that is understandable to humans. This must be guaranteed over conversions. Conversions and other operations on the information must be sufficiently documented to assure possibilities to reconstruct and restore the layout of the screen as it was within the creating authority.

## Analysis

Accuracy refers in the literature to both correctness and reliability of information. Respondents in the two Swedish agencies thought that correctness and reliability as important for preservation criteria. According to definitions in the literature and in the empirical study, correctness and reliability are two different criteria which need to be fulfilled individually in order to fulfill information accurateness. Errors can be the result of preservation activities, e.g., conversion, which may lead to loss or damage of information. Reliability assumes that information content and descriptive information remain reliable over time. Additional descriptive information created by preservation agencies also need to be reliable.

The criterion affiliation with metadata and additional layer of information, which we found in literature, is an important quality criterion for the Swedish agencies. This criterion is desirable on the descriptive level of information and is important to fulfill the completeness, correctness and reliability criteria. It contributes to increased understandability, trustworthiness and usability of information.

Information needs to be trustworthy according to the literature and the empirical study. In long-term digital preservation trustworthiness refers not only to the content information, but also to descriptive information, since it contributes to generation of information meaning. All activities conducted on a preservation object must be documented and traceable through the whole curation process, thus contributing to increased trustworthiness. Users need to believe in and trust procedures and methods used by curators in preservation activities and that preservation agencies providing information are acting in good faith. Belief may originate from the use process. Users form a subjective perception about information quality based on their own judgment, assessment and expectations. Belief, of course can also take its origin from other external factors as recommendations or reputation. Although respondents did not express reputation as a clear quality criterion, they indicated that the reputation of the preservation agency and level of validity play a significant roll in fulfilling the trustworthy. In preservation authorities information needs to be of good or appropriate reputation over time in order to be trusted and used.

In the Swedish agencies information completeness is of great importance for long-term preservation. Both in the literature and for respondents completeness rests on the descriptive level of information. It entails that information should have an appropriate amount of descriptive information in order to be understandable, usable and useful. This includes clear specification of groups of descriptive information and what the descriptions need to contain to make the information complete.

The contextualization criterion found in the literature corresponds to descriptive information (metadata) in the Swedish preservation agencies. However, contextualized information is only one group of descriptive information. Other identified groups are preservation, administrative, juridical and technical. Descriptive information contributes to generation of information meaning and understanding. Curators need to identify descriptions in each description group constituting an appropriate pre-knowledge in order to fulfill different needs of users.

Both in literature and practice it is important that information is understandable. In the literature understandability and interpretability are treated as two separate criteria, while in Swedish preservation agencies understandability also covers interpretability. In this sense understandability involves the cognitive process of the interpretability were balance between the sufficient amount of descriptive information and the appropriate requirement of users' pre-knowledge needs to be reached to get the maximum understanding. An appropriate amount of descriptive information facilitates understandability for some users, but this amount can be insufficient for other. That is why understandability rests partially on a subjective estimation. If information is not understood, it cannot serve as information and therefore cannot be trusted or used.

In the literature the fractional criterion refer to several subcriteria that need to be fulfilled separately. Fractional information need to be quireable and navigable, searchable and manipulative. In the Swedish preservation agencies searchable criterion was discussed, which means implicitly quireable and navigable and manipulative information is undesirable. Curators or users need to be able specify a search to a document, paragraph, table or word levels. This criterion is closely related to the structure criterion, which need to be fulfilled, since users need to navigate along the structures. Words, phrases, paragraphs or a whole document can be used as queries. This implies that the content and descriptive information need to have a structure and need to offer alternative ways to choose between.

Readability is an essential property of information for users. In the literature the emphasis is put on technical contribution, while in the Swedish preservation agencies readability means something that contributes to understanding, usability, usefulness and trustworthiness.

Definition of the criterion of structure in the literature corresponds to the criterion organized in the Swedish preservation agencies. Information need to be structured in such a way that it is understandable, usable and useful. Information structure needs to be explicit and visual to help users in searching and in understanding processes. Otherwise the information will not be understandable readable, usable or useful.

In the literature the validity criterion corresponds to agencies wish to verify that procedures and processed are used accurately in practice. This criterion contributes to trustworthiness of information and usability.

The timeliness criterion was identified both in the literature and the empirical study, however with some different denotations. In the literature timelines refers to the usefulness criterion, while timelines in the Swedish preservation agencies refers to information description level, where documentation about performed preservation activities needs to be up-to-date. This

criterion is important for fulfilling traceability, reliability, trustworthy, completeness and accurateness.

In the literature we found that information needs to be traceable or explicable. This criterion was identified in the empirical study as well. Traceable information is the characteristic of historical sequence of events on the descriptive level of information. The historical sequence of such events includes description of origin and of operations that have been performed on the information. This criterion is important to fulfill in order to increase reliability, trustworthiness, understandability and usability.

Usable information is a central quality criterion for digital preservation, which both the literature and the empirical study showed. Information needs to fulfill several quality criteria in order to be usable: accurate, complete, described, fractional, readable, trustworthy, understandable, valid and traceable.

Usefulness is described in the literature as a criterion of value added for users. This corresponds to useful and sufficient information in the Swedish preservation agencies. It means that useful information is meaningful information. Other criteria found in literature, such as concise and relevant information are also to be included in usefulness criterion as they may provide value to users. Usefulness can be viewed in different ways and is assessed very subjectively. Information can be useful in one context and useless in the other. This criterion deals with user's assessment in obtaining information for different needs. Therefore usefulness doesn't belong to information quality, but rather serve as signal for preservation agencies whether information is of good or poor quality.

The criterion of original state of information is common for the Swedish preservation agencies. However, in the literature original state correspond to the authenticity criterion in archival practice and reliability in library practice. Thus initial state refers to state of information, the state in which it was received by a preservation agency. In practice, information does not need to be identical in order to be understood and used.

In the empirical study we identified a perception criterion, which implies that information needs to be apparent in order for users to be able to read, understand and use information. However, visual state refers only to textual information, which in our view doesn't refer to other human perceptions, like e.g., audio information.

Criteria that we found in the literature but not in the empirical study are coherence, identifiability, objectivity and reputational. But if information is not identifiable, it cannot be found and used and identifiably is therefore essential for preservation even if it was not found in the empirical study.

## Conclusions

The studies have resulted in identification of a number of significant properties that digital information should satisfy in order to retain meaning for users and that are in accordance with requirements of archival and librarian work practice.

The literature review showed that guiding quality principles need to be established in order to ensure the future for preserved digital information. Those guiding principles can be expressed in terms of quality criteria or significant properties for continuous assessment of current state of preserved digital information. They

will also help to identify the metadata set required for preservation of digital information within preservation authorities.

The empirical study at Swedish preservation agencies confirmed that knowledge on how to guarantee quality of preserved digital information is low and that quality work needs to be carried out. The need for a unified information quality assurance framework was also highly confirmed.

We found that technical criteria for data quality are more widely discussed both in the literature and the interviews than criteria for information quality.

## References

- [1] E. Borglund, "RQAM: A Recordkeeping Quality Assessment Model Proposal, *International Journal of Information Quality*, 1, 3, pg. 326 (2007).
- [2] E. Borglund, *Design for Recordkeeping: Areas of Improvement*. (Doctoral Thesis, Mid Sweden University, Sundsvall, 2008).
- [3] Council on Library and Information Resources, *Authenticity in a Digital Environment*. (Washington, D. C.: CLIR, 2000).
- [4] DL.org, Quality Working Group. (2009). <<http://www.dl.org.eu/>>
- [5] C. M. Dollar, *Authentic Electronic Records: Strategies for Long-term Access*. (Cohasset Associates Inc., 2000).
- [6] DPC. Digital Preservation Coalition, *The Preservation Management of Digital Material Handbook*. (2002). <<http://www.dpconline.org/text/intro/definitions.html>>
- [7] L. Duranti, "The Long-term Preservation of Accurate and Authentic Digital Data: The InterPARES Project," *Data Science Journal*, 4, 25, (2005).
- [8] C. J. Fox, A. V. Levitin, T. C. Redman, "The Notion of Data and Its Quality Dimensions," *Information Processing & management*, 30, 1, pg. 9 (1994).
- [9] M. Ge, M. Helfert, "Information Quality Research: Current State and Future Research Agenda," *China Science & Technology Resources Review*, 40, 1, pg. 5 (2009).
- [10] M. Holmes, "The Multiple Dimensions of Information Quality," *Information System Management*, 13, 2, pg. 79 (1996).
- [11] A. Howell, *Manual: Preserving Digital Information – Challenges and Solutions* (2007).
- [12] ICA, International Council on Archives. *Principles and Functional Requirements for Records in Electronic Office Environments. Module 3: Guidelines and Functional Requirements for Records in Business Systems* (2008).
- [13] InSPECT, *Investigating the Significant Properties of Electronic Content over Time*. (Final Report, 2009). <[www.significantproperties.org.uk/inspect](http://www.significantproperties.org.uk/inspect)>
- [14] InterPARES 2, *International Research on Permanent Authentic Records in Electronic Systems*. (Focus Task Force Report, 2008). <<http://www.interpares.org/ip2/book.cfm>>.
- [15] JISC (2008). <[www.jisc.ac.uk/whatwedo/programmes/preservation/2008](http://www.jisc.ac.uk/whatwedo/programmes/preservation/2008)>
- [16] B. K., Kahn, D. M. Strong, and R. Y. Wang, "Information Quality Benchmarks: Product and Service Performance," *Communications of the ACM*, 45, 4 (2002).
- [17] K. Kelton, K. R. Fleischmann, W. A. Wallace, "Trust in Digital Information," *Journal of the American Society for Information Science and Technology*, 59, 3, pg. 363 (2008).
- [18] Y. W. Lee, L. L. Pipino, J. D. Funk, R. Y. Wang, *Journey to Data Quality* (The MIT Press, England, 2006).
- [19] C. L. Liew, S. Foo, *Derivation of Interaction Environment and Information Object Properties for Enhanced Integrated Access and Value-adding to Electronic Documents*, *Aslib Proceedings*, 51, 8, pg. 256. (1999).
- [20] D. Loshin, *Enterprise Knowledge Management: the Data Quality Approach* (San Diego Morgan Kaufmann, 2001).
- [21] J. Nilsson, *Preserving Useful Digital Objects for the Future* (Doctoral Thesis, Luleå University of Technology, Luleå, 2008).
- [22] H. Quisbert, *On Long-term Digital Preservation Information Systems: A Framework and Characteristics for Development* (Doctoral Thesis, Luleå University of Technology, Luleå, 2008).
- [23] H. Quisbert, M. Korenkova, A. Hägerfors, *Towards a definition of digital information preservation object* In ed. M-A Sicilia & M. D. Lytras (*Metadata and Semantics*. Springer US, 2008). <<http://www.springerlink.com/content/j5u68311w2k223q5/>>
- [24] L. L. Pipino, Y. W. Lee, R. Y. Wang, "Data Quality Assessment," *Communications of the ACM*, 45, 4 (2002).
- [25] L. Pipino, R. Wang, D. Kocso, W. Rybolt, *Developing Measurement Scales for Data-Quality Dimensions* (2005).
- [26] B. Reed, *Legal and Recordkeeping Issues Associated with Management of Websites*, *EDPACS, The EDP Audit, Control and Security Newsletter*, XXVIII, 2 (2000).
- [27] M. Runnardotter, H. Quisbert, J. Nilsson, A. Hägerfors, A. Mirjamdotter, *The Information Life Cycle – Issues in Long-term Digital Preservation*, *Proceedings IRIS 28 – Information system search seminar in Scandinavia*. Kristiansand, Norway. (2005).
- [28] M. Runnardotter, *On Organizing for Digital Preservation*. (Doctoral Thesis). Luleå University of Technology. Luleå, 2009).
- [29] R. Ruusalepp, *Digital Preservation in Archives: An Overview of Current Research and Practices*. (A report to the Swedish National Archives. Estonian Business Archives, 2005).
- [30] Y., Wand, R.Y. Wang, "Anchoring Data Quality Dimensions in Ontological Foundations." *Communications of the ACM* 39, 11, pg. 86 (1996).
- [31] R. Y. Wang, D. M. Strong, "Beyond Accuracy: What Data Quality Means to Data Consumers," *Journal of Management Information Systems* 12, 4 (1996).
- [32] R. Y., Wang, E. M., Pierce, S. E. Madnick, G. W. Fisher, (editors) *Information Quality* (Armonk, New York., M. E. Sharpe, Inc., 2005). <<http://site.ebrary.com.proxy.lib.ltu.se>>
- [33] A. Williamson, "Awareness of quality assurance procedures in digital Preservation," *Library Review*, 53, 4, pg. 204 (2004).
- [34] R, K. Yin, *Case Study Research: Design and Methods* (SAGE Publications, 2003).

## Author Biography

*Margarita Korenkova is PhD student in Computer and Systems Science at Luleå University of Technology. Her research is focused on Long-term Digital Preservation.*

*Ann Hägerfors received her PhD in Informatics from Lund University (1994). She is since 2000 a professor in Computer and Systems Science at Luleå University of Technology. Her research is focused on Long-term Digital Preservation. She is the scientific coordinator for the Long-term Digital Preservation Competence Centre.*