Information quality in the context of CRIS and CERIF

Maximilian Stempfhuber
GESIS-IZ Social Science Information Centre, Bonn

Summary

While standardizing on metadata schemas and controlled vocabularies are commonly used measures to improve the quality of a CRIS, there still seems to be only a vague common notion what (information) quality is in detail, which makes it hard to re-use quality assurance processes between CRISs or to implement them in a compatible way, as soon as quality criteria can not be covered by the CERIF data model itself. This paper first tries to summarize the experiences made by the CRIS community to improve the quality of their information systems and then to contrast them with the state of the art in information quality research. Finally, aspects are presented which can stimulate the discussion about a more formalized model of quality in the context of CRIS, which could complement or extend the already existing CERIF standard.

1 Introduction

The quality of information and the quality of services built on this information always have been a major concern of the people and institutions making up the CRIS community. As a community of practice, they share experiences and tools in order to reach a common goal: Design, implementation, use and maintenance of Current Research Information Systems (CRIS).

While quality is a key issue of a CRIS to be useful and to be used, and the need to assess, improve and assure quality has been highlighted in many publications, we feel that an overall, agreed-on concept of information quality, which can be translated into actual systems, is not yet there. Not only is quality defined very differently across CRISs, also much of the efforts put into information quality currently focus on standardization, e.g. by implementing CERIF as a data model, using controlled vocabularies for subject indexing, or checking the structure of harvested data.

Surely, there is nothing wrong with standardization and checking data against good standards, but looking at the problems and solution described in CRIS-related papers during the last four CRIS conferences, an adequate way for approaching quality in the same sense as CERIF supports consistent structure and semantics across CRISs is not available.

What in our opinion is needed is a formal or formalized way to describe quality at the level of the individual data item and at the level of a CRIS as a whole. Formal means that on the one side it generalizes on the rich set of experiences available in the CRIS community (good or even best practices) in a way that it is directly applicable in a variety of contexts and reduces there the risk of failure and inefficiency. Especially with the vision of connecting the European Research Area at the ICT level by networking distributed sources of re-

search information, the notion of quality has to be made much more explicit than it is right now, to support discussion and negotiation at the international level.

This paper will not present a concept for quality in the context of Current Research Information Systems, but we hope to stimulate discussion about such a concept by presenting an overview of the current discourse on information quality in the CRIS community and by pointing to aspects we think are crucial for realizing such a quality model.

2 The Code of Good Practice and Information Quality

The Code of Good Practice (CGP), created by euroCRIS[1], is a document which accumulates the rich experience euroCRIS members made while developing current research information systems at different institutional and regional levels. It guides those intending to create a CRIS through the whole process, or lifecycle, giving advice on how to plan, to realize and to maintain it. Being a generic document claiming to support all kinds of CRISs (e.g. local, regional, international; domain-specific or interdisciplinary) it formulates essential questions which necessarily come up during the lifecycle and which have to be answered in detail to make the effort of creating a CRIS successful.

The central aspect of the CGP certainly is which types of information belong to a CRIS. It defines projects, research programmes, results, publications, organizations etc. as central entities, and the attributes used to describe them together with the relationships between entities and classes of entities make up the conceptual model which should as precisely as possible reflect the same entities from the real world. The conceptual model of a research environment promoted by the CGP is formalized in the CERIF data model (Common European Research Information Format) which is also promoted by euroCRIS.

The view on quality within the CGP can be summarized as *fit for purpose*, one of five views towards quality introduced by Harvey 1995, the others being the *exceptional* view (quality as something special), *perfection* (quality as a consistent or flawless outcome), value for money (quality in terms of return on investment), and *transformation* (quality in terms of change from one state to another). The decision about the *purpose* of a CRIS directly influences most other decisions, like what and how to store, how much information has to be collected, and what features have to be provided so that users can access and process the information in the way they intend to do.

While the *fit for purpose* view to a CRIS surely will be the one that can drive the development of a CRIS throughout most of its lifecycle, it alone can not guaranty the success of a CRIS. The reason in our opinion is not that the CGP is e.g. not comprehensive enough, that some important question one has to ask oneself or the prospective users is missing, that a phase of the lifecycle is not covered, or that a necessary feature for end users is not mentioned. The reason is – like it is the case with many guidelines for user interface design – that despite its richness of details it is too abstract to be directly applied in a deterministic way.

For many of the advices given in the CGP, e.g. concerning quality assurance or end user features, there is an unlimited number of ways criteria can be defined or features can be designed and implemented. It does not restrict enough the degree of freedom in defining quality in a CRIS so that it would be comparable between individual systems – a central require-

1 http://www.eurocris.org

ment for connecting distributed CRISs and building services on it which themselves match predefined quality criteria and are fit for purpose.

Much of the success – and quality – of current CRISs surely can be attributed to the comprehensiveness of the CGP (and, of course, the CERIF data model) and the high level of expertise the makers of CRISs are able to contribute during the development of these systems. But we feel that the aspect of information quality in the context of CRISs needs more formalisation to deliver the same level of guidance the CERIF data model already provides for data structures and semantics.

3 The CRIS Community's Approach to Quality

The quality of current research information systems – especially the quality of their contents – has always been an issue for the CRIS community. In the following section we try to give an overview of some aspects of quality, which have been raised by a number of authors at the CRIS conferences between 2002 and 2008. The proceedings of these conferences have been scanned cover-to-cover for articles either explicitly addressing aspects of quality in their title or dedicating one or more sections to this topic[2]. Furthermore, the text of all articles potentially raising aspects of quality in the context of specific CRISs, harvesting data, promoting certain metadata schemas or using CRIS for research evaluation (where the quality of data is essential) has been scanned for individual paragraphs on quality issues related with research information itself, the process involved to create this data or requirements on the use of the data. In total, 69 articles from CRIS conference proceedings have been analyzed (poster and workshop abstracts being omitted) from which 16 articles have been used for this overview.

3.1 Proceedings of CRIS 2002

In their paper "CERIF – Information Retrieval of Research Information in a Distributed Heterogeneous Environment", Lopatenko et al. 2002 describe three different approaches for distributed information retrieval in CERIF-compliant CRISs: Distributed databases, Semantic Web and Web Services. Concerning the discussed technologies for querying distributed CRISs they conclude that the CERIF data model provides formal correctness and the needed flexibility to support these technologies and to be flexible enough for futures demands towards current research information systems. In regard to data quality they state: "We believe that main parameters of quality of scientific data are completeness, actuality and correctness. Completeness of data in CRIS is presence of data about all entities which are subject of given CRIS." While in regard to one specific CRIS they see administrative efforts for data harvesting, analysis and input as preconditions for correctness, actuality and completeness, in distributed environments reasoning procedures are required in cases quality-critical applications are using the collected data. While – at this time – actuality and correctness have not been addressed by their distributed CERIF project, metadata about completeness was introduced using DARPA Agent Markup Language (DAML[3]) and proposed as an extension to CERIF.

2 The proceedings of CRIS conferences do not contain keyword indexing at this time, so no filter criteria could be defined to select relevant articles for this analysis in a more formal way.

3 http://ww.daml.org

The aspect of trust and quality of service is mentioned by Matthews 2002 in his article "Integration via Meaning: Using the Semantic Web to deliver Web Services". The author views both technologies as complementary, with Web Services providing a loosely coupled middleware solution and the Semantic Web providing means for formal knowledge representation, logic and reasoning. The Semantic Web in particular also could provide functions for establishing trust by adding access policies to information resources encapsulated by Web Services, user agents could present credentials allowing them to access these resources, and via negotiations the suitable access rights and obligations for data providers and users could be negotiated. Also the quality of service in terms of response time of such an information service, the accuracy of the data delivered and the level of confidentiality could be expressed in Semantic Web terms and negotiated by the involved parties.

Lepori&Cantoni 2002 approach quality of CRIS from the user's point of view, specifically referring to the relationship a CRIS can build with the target audience it is designed for (in contrast to publishing information to "everybody"): "Is there any user for this CRIS?" They start with the finding that Internet communication is rapidly evolving from a model where information is published for a generic audience and where quantity and quality are considered the most important factors towards a personalized publishing model, trying to establish a long-term relationship with the audience: "The major implication is that to be a successful provider on the Internet it is not anymore sufficient to publish large amount of information, even of very good quality, but one needs to draw from this information very specific services tailored to the needs of specific user groups; moreover, these services have to be better than those which can be obtained through other communication." They stress that it is not sufficient to just improve the data collection by applying standards like CERIF or using controlled vocabularies to index them, but these activities have to go side-by-side with generating useful services for dedicated groups of CRIS users.

Pretty much in line with Lepori&Cantoni 2002 are the findings of Koopmans 2002. In her article "What's your question? The need for research information from the perspective of different user groups" she presents the results of a field study where the main questions tried to find out what kind of research information users need, what information resources were used and which are missing at the moment. To make research information accessible to user groups with different levels of knowledge she suggests providing different information layers with different levels of detail and accessibility, and to enable browsing in data collections by means of taxonomies or thesauri which should be developed in co-operation with the specific user groups. Again, certain aspects of data quality (e.g. controlled vocabularies) are seen as enablers of better quality of service at higher levels of a CRIS.

3.2 Proceedings of CRIS 2004

Aspects of quality were mentioned in Covey 2004, "Global Cooperation for Global Access: The Million Book Project". The paper describs a world-wide initiative to make the full text of books freely available on the Web, i.e. as parts of Digital Libraries which might complement current research information systems in providing direct access to publications which are referred to from publication entries in a CRIS. Quality in this context is primarily dedicated to produce correct bibliographic metadata for the scanned books and for optical character recognition (OCR) to convert the scanned image of each page into machine readable text. The quality of both processes directly influences the accessibility of the digitized books, especially in situations where metadata or digitized materials are exchanged be-

tween (search) service providers (metadata and file formats have to adhere to standards) or where full text search in the digitized books is offered to users.

3.3 Proceedings of CRIS 2006

Research evaluation and benchmarking was the context in which Hornbostel 2006 presented his keynote entitled "From CRIS to CRIS: Integration and Interoperability". He stressed that distributed CRISs could support the more and more complex environment and interrelations between information for the research process, evaluation for (public) reporting and benchmarking for control processes. Not only have they to be transformed from local information systems to networked systems forming a distributed information environment, they have to become (permanent) monitoring systems, for which not only bibliographic information but also reliable data on research institutions is required. The overall quality of a CRIS, the author states, depends on the benefits researchers can draw from them in providing, accessing and using research data. Individual aspects of CRIS quality are making work easier (e.g. by automating processes or making them more efficient), avoiding unnecessary or duplicate work (e.g. by link up a CRIS with external systems) and – last but not least – by tightly incorporating CRISs within research administration, decision making, evaluation and reporting. Standards for structuring data, like CERIF, are a precondition for this as they provide a high level of semantic and technical interoperability, but they must not only be followed locally but also on the international level.

Standardization is also of major importance with the research funding databases of the Deutsche Forschungsgemeinschaft, DFG. For "The Use of Research Funding Databases for Research Assessment Information Systems" (Bovelet 2006) it is essential to correlate information from multiple, heterogeneous sources into a unified model, which is in this case been done by compiling a concordance between the different subject classifications and the records of research institutions from multiple sources. This process requires direct involvement of domain experts (e.g. officers or review committees) to yield a quality level of this mapping which is high enough to produce reliable statistics across this heterogeneous data. But the efforts used in this process and the improvement in data quality generate many benefits for funding agencies, especially for new information services and management purposes (Güdler 2006).

Typical problems with information quality in "IWETO, a research information clearinghouse" for the Flemish University Council (van Grootel 2006) led to a re-design in information flow and overall processes (the "Research and Innovation Value Chain") avoiding the problems and resources needed for ex-post quality assurance. In this case, minimal agreed quality criteria for primary business objects (e.g. project, research team, researcher) were not adhered to at the stage of decentralized data input, but were only detected at a later point in time, when information was collected from the multiple sources and integrated into a centralized system. In the re-designed model of IWETO, the enforcement of standards to improve data quality is delegated to the applications and workflows making up IWETO at the process and functionality level. Jeffery&Asserson 2006 support this process model in that they suggest that data is entered into a CERIF-compliant CRIS and verified at the earliest possible stage, which allows newly entered data to be checked for consistence against already available data. Also reusing existing data for suggesting values during data entry can help to improve data quality at the time of input. And by using logic-properties of the CERIF data model, data values can be constrained to improve quality.

In the context of research assessment, access to the complete publications of the entity under evaluation is essential. Hey et al. 2006 state in their article "Leveraging the Institutional Research Repository: harnessing the drive for quality assessment", that the quality of repository contents is crucial and – besides bibliographic data – can be improved by add-ons like editorial reviews and persistent identifiers, which support long-term availability and consistency of publications and their linkage.

3.4 Proceedings of CRIS 2008

To support "Quality Assurance in the Research Documentation System Frida" (Lingjærde& Sjøgren 2008), authoritative registers (e.g. for periodicals, publishers and institutions) which are shared by the participating institutions are used to improve consistency and quality through standardization. For the import of publication data, the model of authoritative data is also used: Data from specific, authoritative sources will overwrite manually entered data, and certain attributes of authoritative data can not be manually changed afterwards to guaranty correctness. Via quality reports, inconsistencies in the data can be detected, especially across the different institutional databases connected to FRIDA.

An iterative process for quality enhancement of the research information available at the German Research Foundation (DFG) is presented in the paper "Quality is the Product is the Quality. Information Management as a Closed-loop Process" (Güdler&Hahnen 2008). The basic concept employed here is that users of information products can directly provide feedback on omissions or incorrect data, which helps to improve the underlying database as a whole whenever parts of it are actually used (and corrected in the case of errors).

In the context of the "Enhancement of Data Quality in Distributed Open Access Repositories", Severiens et al 2008 present the DINI certificate for repositories, which e.g. defines repository content by collection policies, together with additional measures to improve metadata quality at the level of the backbone of the proposed OA-Network (e.g. by validation and plausibility checks, de-duplication and by crosswalks between heterogeneous subject classifications and DDC).

Digital Author Identification (DAI) is a measure for improving information quality at the national level in the Netherlands, with the goal of "Creating an Academic Information Domain" (Baars et al. 2008). The vision is, that by assigning a DAI to all researches in the Netherlands harvested metadata for publications, research data etc. (which will contain DAIs) will be easier to link and integrate.

In a case where information is harvested but Digital Author Identification or other persistent identifiers are not available, automatic means for data integration – and therefore quality enhancement – have to be applied. In their paper "Analyzing European Research Competencies in IST: Results from a European SSA project" (Jörg et al. 2008) the authors present a solution for de-duplication based on machine learning.

3.5 Summing up

Based on this literature review, four main areas can be defined which currently seem to be of central interest to the CRIS community:

- Information quality: To improve on aspects like correctness, authoritative registers, controlled vocabularies, persistent identifiers, automatic checking of values and structure are used, and through intellectual processes carried out by experts the data is enriched to

make it more useful or trustworthy. Semantic Web technologies are suggested to improve completeness of data (also across individual CRISs).

- Data integration: This becomes an issue as soon as data is exchanged or individual CRIS are networked. Methods employed are the certification of information systems, checking of data structures and values against formal requirements, mapping between vocabularies, and automatic and intellectual de-duplication.
- Quality as a process: Checking data towards quality criteria as soon as it is created, using existing data to verify new data, and enabling feedback loops from users of data to incrementally improve overall data quality.
- Personalization: Better matching CRIS features (e.g. amount and level of detail of data, presentation of information, availability of features) to the specific demand of individual users or well defined user groups.

4 A Closer Look to Information Quality

To support the discussion on quality in the context of CRISs, a closer look at the state of the art in information quality research is helpful. This is not meant as an in-depth or complete overview, but to list essential requirements towards a more explicit quality concept for CRISs.

Much of the current research on Information Quality (IQ) is based on the work of Wang et al. 1993 and Strong et al. 1997a and 1997b. Their research was motivated by IQ problems and IQ management errors which occur in all types of organizations, and where financial losses and failed business initiatives are the results – and sometimes also casualties. Two streams of research which try to prevent IQ problems are IQ assessment and IQ management, and one of the central conferences where the IQ research community meets is the International Conference on Information Quality[4] (ICIQ), annually held since 1996 at the Massachusetts Institute of Technology (MIT).

The concept of IQ is adapted from the general concept of quality and can be defined from an information consumer perspective and a data perspective (Ge&Helfert 2007). From the consumer perspective, IQ can be defined as information that is fit for use, a definition which is problematic since users are not very capable of finding errors or adapting their information use to the erroneous data. From the data side, IQ can be defined as information meeting specifications or requirements, which is much easier to formally express and assess.

Ge&Helfert 2007 provide a framework for IQ assessment review (see) which covers on its three layers the essential concepts and their relationships. They define IQ assessment "as the process of assigning numerical or categorical values to IQ dimensions in a given setting". At the lowest level (Metric Layer) one can find IQ metrics which represent the different IQ problems identified for a domain, which in turn are arranged in a "2 context views x 2 assessment views" model.

Context independent IQ problems from a data perspective e.g. are spelling errors, incorrect values, outdated data or wrong text formatting, whereas from a user perspective they are e.g. inaccessible or insecure information or information that is difficult to aggregate or transform. Context-dependent IQ problems from a data perspective are the violation of domain constraints or the violation of company or government regulations, whereas from a

4 http://mitiq.mit.edu/ICIQ/

user perspective they are e.g. information not based on facts, information that presents an impartial view, is hard to manipulate or hard to understand.

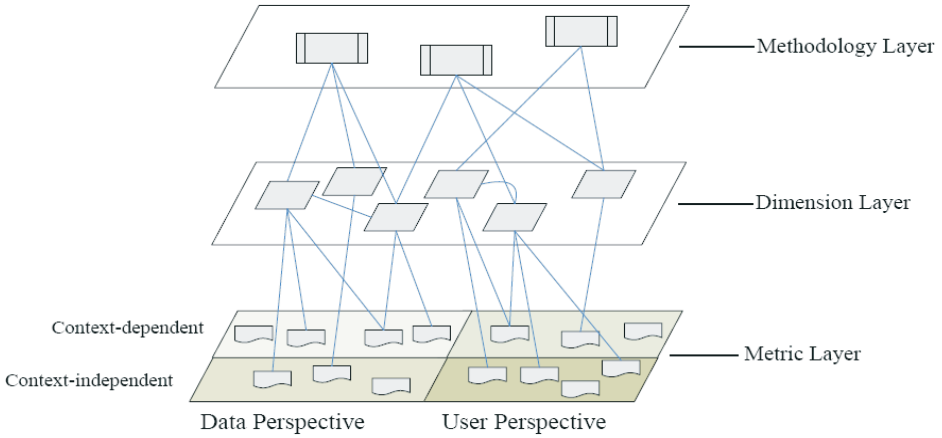

Figure 1: Framework for IQ assessment review (Ge&Helfert 2007)

The Dimension Layer consists of the different IQ dimensions which are characteristics of the information, like accuracy, timeliness or completeness. These IQ dimensions are linked in a many-to-many relation to IQ metrics, so that the IQ dimension of accuracy can link to the IQ metrics incorrect data and out-of-date data; the IQ metric out-of-date-data can link to the IQ dimensions accuracy and timeliness. In cases where one IQ metric links to more than one IQ dimension, a relationship between these IQ dimensions is established and changes in the IQ metric will influence both IQ dimensions, which can also lead to a decline of one IQ dimension as soon as some other is improved (negative correlation between IQ dimensions). Much research has been done in relevant IQ dimensions, around 180 different dimensions can be found in the literature (cf. Naumann&Sattler 2006).

At the Methodology Layer, several different approaches to IQ assessment exist, one of which is the product and service performance model for information quality (PSP/IQ, cf. Kahn et al. 2002). It maps IQ dimensions onto four quadrants in the PSP/IQ model, comprising product and service quality in one dimension, and conformance to specifications and meeting consumer expectations on the other (figure 2):

- Product quality/Conforms to specifications: Indicates that the information meets standards of accuracy, completeness, and free-from-error (sound information).
- Service quality/Conforms to specifications: Indicates a process by which information consumers regularly receive information in a timely manner (dependable information).
- Product quality/Meets consumer expectations: Information product must be useful and relevant to the user's needs (useful information).
- Service quality/Meets consumer expectations: Information consumers can easily obtain and manipulate information that adds value to their task (usable information).

	Conforms to Specifications	Meets or Exceeds Consumer Expectations
Product Quality	Sound Information • Free-of-Error • Concise Representation • Completeness • Consistent Representation	Useful Information • Appropriate Amount • Relevancy • Understandability • *Interpretability* • *Objectivity*
Service Quality	Dependable Information • Timeliness • Security	Usable Information • Believability • Accessibility • Ease of Manipulation • Reputation • Value-Added

Figure 2: Mapping the IQ dimensions into the PSP/IQ model (Kahn et al. 2002)

In addition to the research presented here, much research has been done in actually assessing IQ with questionnaires, by measuring and profiling and by aggregating single IQ scores, and in IQ interpretation, e.g. in annotating query results with IQ scores for ranking, for selecting information sources in regard to their IQ scores, or as a cost parameter in query optimization.

A last area to mention is IQ improvement, e.g. by data cleaning, which may involve profiling, normalization/standardization, detecting missing values and detecting outliers (cf. Naumann&Sattler 2006). Especially in the context of (relational) databases and data warehouses, sophisticated software is available for this purpose.

While not exhaustive at all, this section should give an overview on IQ and IQ assessment, and an idea of the models and formal methodologies already available in this area. We feel that the aspect of quality in current research information systems could greatly benefit from adapting some of the views and approaches presented and that this should lead to an extension or complement to the CERIF standard and/or the Code of Good Practice.

5 Connecting Information Quality and CRIS More Explicitly

Looking at the current developments in the area of current research information systems, like the requirement to network individual, often heterogeneous CRISs or to provide reliable information for research evaluation at regular intervals, advancement in information quality increases a CRIS' fitness for use. We therefore suggest the following aspects as relevant to be discussed in the CRIS community with the goal of extending or complementing our current standards and practices:

- The lifecycle for CRISs proposed in the Code of Good Practice is currently following the waterfall model. Research in software technology and experiences from CRIS projects (cf. Güdler&Hahnen 2008) suggest that an iterative model with loopback cycles could improve both, system development and information quality.

- The view of "quality as a process" (cf. van Grootel 2006 and Jeffery&Asserson 2006) suggests that IQ assessment must be consistently applied to all business processes within a CRIS. Furthermore, we argue that also in the CRIS lifecycle IQ can not be covered by a single phase, but a consistent model of IQ assessment has to be applied all the way from concept creation to regular operation of a CRIS.
- The "fitness for use (or purpose)" view to IQ and the selection of only a few IQ dimensions, like in the current version of the CGP, might in many cases not be sufficient to guaranty the required quality of service. What is needed is an overall model of IQ in the context of CRISs with a complementary set of IQ metrics and IQ dimensions which can detect typical IQ problems.
- To support personalization as indicated by Lepori&Cantoni 2002 and Koopmans 2002, matched sets of information needs, CRIS (and data) features, and ways for implementation should be created to lead the design of a CRIS and to satisfy user expectations as a relevant dimension of information quality. Standardized questionnaires could help in assessing user expectations in a consistent way.
- The IQ features of a CRIS should be formalized in a human and a machine-readable way to support automatic (and better) source selection, information retrieval, ranking and processing of results, not only during user interaction with a specific CRIS but also in distributed environments. This could be done as a collection level description (CLD) extended with IQ attributes and by extending CERIF with IQ attributes at the entity or attribute level.
- Best practice in IQ improvement, e.g. in data cleaning or data enrichment with persistent identifiers, should be modularized, documented and made publicly available for re-use.

6 Conclusions

Quality in the context of CRISs has been an issue since the first version of the Code of Good Practice was drafted. Literature review in the proceedings of the last four CRIS conferences showed that a large number of publications mention different aspects of – or problems with –information quality. IQ is crucial for acceptance of information systems by users and as a prerequisite to build services adhering to a predetermined quality level.

Parallel to the development of the CGP and the CERIF data model during the last 6 years, numerous advances have been made in information quality research. Combining results from quality research with the rich experience of the CRIS community in handling quality problems – especially in settings with distributed and heterogeneous information – could lead to a more formal and unified view to IQ, which could foster discussion and knowledge exchange in the CRIS community on one hand and ease the transfer of standardized IQ metrics, dimensions and methodologies between CRISs on the other. Not only would the producers and users of CRISs gain from such a development, also the European Research Area would benefit from higher information quality and easier exchange of research information across regions and disciplines.

7 References

Adamczak, W.; Nase, A. (eds.) (2002): *Gaining Insight from Research Information. 6th International Conference on Current Research Information Systems.* Kassel: Kassel University Press.

Asserson, A.; Simons, E. J. (eds.) (2006): *Enabling Interaction and Quality: Beyond the Hanseatic League. 8th International Conference on Current Research Information Systems.* Leuven: Leuven University Press.

Baars, Ch.; Dijk, E.; Hogenaar, A.; van Meel, M. (2008): *Creating an Academic Information Domain.* In: Bošnjak&Stempfhuber 2008, pp. 75-85.

Bošnjak, A.; Stempfhuber, M. (eds.) (2008): *Get the Good CRIS Going: Ensuring Quality of Service for the User in the ERA. 9th International Conference on Current Research Information Systems.* Maribor: Institute of Information Science.

Bovelet, D. (2006): *The Use of Research Funding Databases for Research Assessment Information Systems.* In: Asserson&Simons 2006, pp. 41-54.

Covey, D. (2004): *Global Cooperation for Global Access: The Million Book Project.* In: Nase&van Grootel 2004, pp. 139-151.

Ge, M.; Helfert, M. (2007): A Review of Information Quality Research - Develop a Research Agenda. *Proceedings of the 2007 International Conference on Information Quality (ICIQ).* (http://mitiq.mit.edu/iciq/ICIQ/iqdownload.aspx?ICIQYear=2007&File=A+REVIEW+OF+INFORMATION+QUALITY+RESEARCH.pdf)

Güdler, J. (2006): *CRISs as a Management Tool: Experiences and Perspectives at the German Research Foundation (DFG).* In: Asserson&Simons 2006, pp. 41-50.

Güdler, J.; Hahnen, H. (2008): *Quality is the Product is the Quality. Information Management as a Closed-loop Process.* In: Bošnjak&Stempfhuber 2008, pp. 32-40.

Harvey, L. (1995): Editorial: The quality agenda. *Quality in Higher Education*, 1 (1), pp. 5–12.

Hey, J.; Carr, L.; Simpson, P. (2006): *Leveraging the Institutional Research Repository: harnessing the drive for quality assessment.* In: Asserson&Simons 2006, pp. 89-98.

Hornbostel, S. (2006): *From CRIS to CRIS: Integration and Interoperability.* In: Asserson&Simons 2006, pp. 29-38.

Jeffery, K.; Asserson, A. (2006): *Supporting the Research Process with a CRIS.* In: Asserson& Simons 2006, pp. 121-130.

Jörg, B.; Ferlež, J.; Uszkoreit, H.; Jermol, M. (2008): *Analyzing European Research Competencies in IST: Results from a European SSA project.* In: Bošnjak&Stempfhuber 2008, pp. 105-121.

Kahn, B. K.; Strong, D. M.; Wang, R. Y. (2002): Information Quality Benchmarks: Product and Service Performance. *Communcations of the ACM.* April 2002, 45 (4), pp. 184-192.

Koopmans, N. (2002): *What's your question? The need for research information from the perspective of different user groups.* In: Adamczak&Nase 2002, pp. 183-192.

Lepori, B.; Cantoni, L. (2002): *Is there any user for this CRIS?* In: Adamczak&Nase 2002, pp. 173-181.

Lingjærde, G.; Sjøgren, A. (2008): *Quality Assurance in the Research Documentation System Frida.* In: Bošnjak&Stempfhuber 2008, pp. 21-27.

Lopatenko, A.; Asserson, A.; Jeffery, K. (2002): *CERIF – Information Retrieval of Research Information in a Distributed Heterogeneous Environment.* In: Adamczak&Nase 2002, pp. 59-68.

Nase, A.; van Grootel, G. (eds.) (2004): *Putting the Sparkle in the Knowledge Society. 7th International Conference on Current Research Information Systems.* Leuven: Leuven University Press.

Naumann, F.; Sattler, K.-U. (2006): *Information Quality: Fundamentals, Techniques, and Use.* EDBT Tutorial, 10. International Conference on Extending Database Technology, Munich, March 28, 2006 (http://www.edbt2006.de/edbt-share/IQ-Tutorial.pdf)

Matthews, B. (2002): *Integration via Meaning: Using the Semantic Web to deliver Web Services.* In: Adamczak&Nase 2002, pp. 159-167.

Severiens, Th.; Klatt-Kaufmann, M.; Malitz, R. (2008): *Enhancement of Data Quality in Distributed Open Access Repositories.* In: Bošnjak&Stempfhuber 2008, pp. 71-75.

Strong, D. M.; Lee, Y. W.; Wang, R. Y. (1997a): Data Quality in Context. *Communications of the ACM*. May 1997, 40(5), pp. 103-110.

Strong, D. M.; Lee, Y. W.; Wang, R. Y. (1997b):10 Potholes in the Road to Information Quality. *Computer*. August 1997, 30(8), pp. 38-46.

Van Grotel, G. (2006): *IWETO, a research information clearinghouse.* In: Asserson&Simons 2006, pp. 69-76.

Wang, R. J.; Kon, H.; Madnick, S. (1993): Data Quality Requirements Analysis and Modeling. *Proceedings of the Ninth International Conference of Data Engineering*, April 1993. pp. 670-677.

8 Contact Information

Dr. Maximilian Stempfhuber
GESIS-IZ Social Science Information Centre
Lennéstr. 30
53113 Bonn
Germany

e-mail: max.stempfhuber@gesis.org
http://www.gesis.org/IZ/Stempfhuber/