



QUALITY CONTROL IN WEB CATALOGUES – CONSIDERATIONS WITH RESPECT TO THE SUBJECT GATEWAY SOCIOGUIDE

Jana Meichsner

*GESIS – Leibniz Institute for the Social Sciences, Schiffbauerdamm 19, 10117 Berlin, Germany, e-mail:
jana.meichsner@gesis.org*

H. Peter Ohly

*GESIS – Leibniz Institute for the Social Sciences, Dreizehnmorgenweg 42, 53175 Bonn, Germany, e-
mail: peter.ohly@gesis.org*

ABSTRACT

Quality is a crucial factor for ensuring acceptance of information services among different user groups. The steadily growing amount of data incorporated in the GESIS SocioGuide can not be managed solely by intellectual means, but relying only on automatic evaluation is often not what is desired of a high standard. In order to meet the diverse requirements (high quality and a multitude of resources) we choose a strategy of semi-automatic quality control. Here we describe some principles and routines of guaranteeing documentary quality in the application field of social science with respect to the subject gateway GESIS SocioGuide based on DBClear system software. Practical components of this quality assurance workflow are: Searching new data, editorial qualitative vs. automated validation, suggestions by users, decentralized workflow.

KEYWORDS

Subject gateway, evaluation guidelines, selection criteria, information quality, quality control, Internet resources



INTRODUCTION

Information is knowledge which is available at the right time and right place for problem-solving. Wersig (1973) speaks of the "reduction of uncertainty," i.e. compared to a previous state of indecision, the user may experience a state of more than random deciding. Similarly, Kuhlen (1995) locates information as "knowledge in action" between data and knowledge, i.e. data are interpretable and are not just part of a knowledge set, but are also an application entity of special interest. Information in the proper sense can only be spoken of when the data actually provided for the information users has a current, specific value. Information societies must also be qualitatively secured, which entails a demand for protection and an increase in information quality (Wiethaus, 2001). The goal of any specialized information supply, including technically provided information must be to ensure this. Eppler (2006) writes in his introduction to the book *"Managing information quality:"* *"Information quality is a term that is vague and general, yet promising and pertinent: Amidst the increasing quantity of information available, the quality of information becomes a crucial factor for the effectiveness of organizations and individuals."* Here the development of standards is of special importance, since information can be judged by the (uninformed) customer only after receipt and never exhaustively (c.f. Rösch/Weisbrod, 2004). In the early days of the Internet, the quantity of documents found was critical, but now with its rapid growth (e.g. Rüter, 2008) there is a demand for restriction to those documents that most precisely match a given question or information request.

2. PORTALS AND THE NEED FOR QUALITY

2.1. Quality Criteria in the Web

Quality of Information

Quality, according to Wiethaus (2001) is a relative property which is affected by material and sociological circumstances. How it is determined depends upon, among other things, the development of technological standards; which, of course, vary over time.

The DIN ISO 8402 standard defines quality as *"the totality of qualities and characteristics of a product or service, which in turn affects their ability to meet fixed or assumed anticipated needs "* (Rittberger, 2004). In connection with information services quality may be seen as a distinguishing characteristic (such as scientific, practical), but also as a value scale (cf. Enderle, 2001). Both dimensions of quality are dependent on the needs of users, since quality can only result through the participation in the process of information transmission (see Hobohm, 1998).



While traditional information agencies accrue confidence via their specialization and institutionalization, on the World Wide Web, where there are many technically, highly capable information providers, the quality of the information supplied is in question. Thus, since the existence of the Web attempts have been undertaken to give (only) a qualitatively filtered offer to the respective users. Search engines already filter a fraction of the sites existing in the Internet using formal criteria before further indexing (visible Web vs. invisible Web). In particular, search engines like Yahoo or Lycos also took advantage of user proposals and editor evaluations. Automatic procedures, like evaluation of word frequencies or number of linking pages have been the usual automatic procedures to weight the universe of Web contents since a long time. However, content quality, and quality in terms of user friendliness is desired, which are inseparable for users.

Huang et al. (1999 (after Mandl, 2005)) more precisely differentiate the following dimensions of information quality (see Table 1).

Table 1: Categories of Information Quality

IQ category	IQ dimensions
Intrinsic	Accuracy, objectivity, believability, reputation
Contextual	Relevancy, value-added, timeliness, completeness, amount of information
Representational	Interpretability, ease of understanding, concise representation, consistent representation
Accessibility	Access, security

According to Huang et al., Mandl emphasizes the importance of context in the user assessment of value, thus being a subjective component (expectation oriented reference). By contrast, the substantive, authoritative component (manufacturer-process orientated reference) should be represented by absolute criteria, as complete, trustworthy, reliable, balanced, and well represented. It becomes clear that even with a substantive point of view, no absolute assessment can be made, because it would require a perfect overview of the scientific field and the website supply. Huang et al. also stress the intrinsic aspect of the accessibility of content in addition to the quality of the content. The structuring of the information offered determines the possibility of navigation and finding the desired information. Another aspect is topicality (see above). On the one hand, it is important that information is also secured generally and for the longer term (at least during the data availability). On the other hand the user only receives an information advantage through news not already generally well-known. The current emphasis may depend on the action context of the user, which can be of a more politico-strategic or more systematic-sustainable nature. At the same time, the information must be presented in a way that conforms to expectations, such that the

user can also process the received information adequately in his context (see Rittberger, 2004). In the end, the user will also attach very personal expectations for utilization to the information received, which can be seen rather as side effects in the given context, and can hardly be guaranteed by the information supplier. These include. prestige, contacts, research means, career and others (cf. Ohly, 2008).

Similarly, quality considerations applying to documents found in the Internet, are also valid for the quality of Internet information brokerage, whether it be the kind of description that is given to the source or the scale applied to the selection of documented Web pages.

Thus quality selection criteria for Internet subject gateways as determined in the DESIRE project are (after Hobohm, 2003):

- Scope (coverage, access, cataloguing)
- Content (validity, topicality, substance, uniqueness, accuracy)
- Form (navigation, support, standards, technology, aesthetics)
- Process (system integrity, coherency)
- Collection Management Policy

The importance of observing quality standards for clearinghouse offers is demonstrated by the IFLA checklist, which discussed several Web-clearinghouses in a joint workshop (see Table 2).

Table 2: Jakob Andersen: IFLA Guidelines for evaluating gateways (presented at the 66th IFLA Council and General Conference Jerusalem, Israel, 3-18 August 2000)

Overall 'feel' Do you like the gateway? ...
Purpose What is the stated purpose of the gateway? ...
Audience Which user group(s) is(are) catered to? ...
Services Apart from providing entrance points to the net, which added services are provided? ...
Scope How does the gateway's breadth (aspects covered), depth (level of detail), and time coverage fit with the stated aims?
Content Are the links/resources provided accurate, authoritative and current?
User interfaces What are your feelings about the graphic/textual design? Does the amount of graphics significantly slow down the transfer of screens?...
Recommendations Do you have any recommendations for the managers of the gateway as to improvements – general or specific?



Subject Gateways as Quality Control

The development of the Internet in the mid-90s led to the creation of information brokerage services (clearinghouses), which give high-quality subject access to scientific sources. No further distinction between clearinghouse, gateway, Internet catalogue and virtual library will be made, particularly since such terms often correspond to the genesis of the respective Internet offering and the services have long since moved on in what they offer, all the way up to portals. Services deserving a closer look include those:

- Accessible in the Web and referring to sources of the Web,
- Leading to other technical service of this kind,
- Created with scientific background,
- Of a qualitative selection or processing,
- And if so, offering any further professional services at the same time.

Services exhibiting only some partial aspects are also considered in the following (cf. Ohly, 2004).

As in Bargheer (2002), we think of subject gateways as general information services which enable professionally focused access to systematically catalogued Internet resources. The Follet Report (1993), for example, initiated the rise of digital libraries in Great Britain. In this context, beginning in 1995 the eLib- programs were implemented, which also included a component for intellectually indexing Internet sites. The EU-funded 'first-generation gateway-related projects' DESIRE (1998-2000) and RENARDUS (2000-2002), which laid the foundations in this area, have to be mentioned in the wider context of indexing the Internet as a scientific publication and communication medium.

Thus, on the basis of a common concept, different European information services were established in the framework of DESIRE, comprising multidisciplinary national Internet resources (e.g. DutchESS, <http://www.kb.nl/dutchess/>) or international resources (e.g. BUBL, <http://bubl.ac.uk/LINK/linkbrowse.cfm/e/>) or resources to specific areas (e.g. SOSIG (today: Intute: Social Sciences, <http://www.intute.ac.uk/socialsciences/>), OMNI (today: Intute: Health and Life Sciences, <http://www.intute.ac.uk/healthandlifesciences/medicine/>)).

In addition, the German Council of Science and Humanities made a demand for: *"Virtual specialized Libraries, which (...) are to be established in the special subject collections libraries, should enable access to relevant scientific digital documents via the Internet"* (after Wissenschaftsrat, 2001).

The term "Quality Controlled Subject Gateways" (QCSG) was shaped in the context of the Desire project (Koch, 2000). In the project description (DESIRE, 1996) this is explained as follows: *"Selective subject gateways on the Internet are characterized by their quality control. The core activities of resource selection and description rely on skilled human input (by Librarians, academics and experts) and are not activities that*



lend themselves to automation. (...) Methods and tools (... Have to be ...) created to assist the staff of subject gateways to develop and maintain their quality control systems:

- *A generic conceptual model is provided, which describes (...) the quality issues (...) relating to each part of the (gateway) process.*
- *A comprehensive list of selection criteria is given, which incorporate 'tips' and 'hints' for evaluating Internet resources ...”*

Even the German Council of Science and Humanities (Wissenschaftsrat, 2001) speaks of "high quality and valued resources." The following criteria for subject gateways are mentioned by Rösch/ Weisbrod (2004) among others:

- contained sources are of a level sufficient for scientific requirements,
- fixed criteria for evaluation and quality control.

Essential tools for quality assurance are the standardized description of the documents by means of metadata from thesauri or (specialized) classifications and the summary of the contents, which are used to characterize the resources (value-added service). Thereby the user is given the possibility to meet a closer selection according to self-chosen combinations. Furthermore, the application of international and technical standards supports the interoperability of the subject gateways, thus, access to larger data sets.

2.2. Quality Guidelines

Information Quality

DIN ISO 8402 standard defines quality as *"the totality of qualities and characteristics of a product or service ..."* (see above; see Rittberger, 2004).

Quality is thus necessarily not an abstract characteristic of a service or product, it must be more closely determined in a certain context. Quality is assumed to be composed of certain characteristics, specific for the appropriate product or service. Wiethaus (2001) stresses that according to social differentiation there is differentiation in the concept of quality and thus numerous, parallel entitled, "applied qualities" (see also El-Menouar, 2002).

What we may essentially understand as information quality improvement depends decisively on whether information is considered as a product or a service, thus production process-oriented or expectation-oriented related. The definition of Kahn *"Quality information is information that meets specifications or requirements"* (See Kahn/Strong, 1998, after Eppler et al., 2000) emphasizes the process view, while *"Information quality can be defined as information that is fit for use"* by Huang et al.



(1999, after Nohr, 2001) emphasizes the service perspective. The evaluation of the quality of information in the information process can now be targeted at two points:

- the selection of the material which should be documented and
- the evaluative description of the material finally documented.

Usually, both criteria are discussed together and not precisely separated. Also, unlike earlier attempts, gateways today rarely give explicit ratings in descriptions of information sources, as the Argus Clearinghouse for Subject-Oriented Internet Resource Guides (see Figure 1) did earlier.

The sources included in the Argus Clearinghouse were evaluated by experts, using five criteria: objective and subjective information content, design, schemes used, meta-information; and special design and organization were also evaluated (see Ohly, 1998). A further aspect of a source's improvement in quality, and, respectively, the support of better filtering of a source is the multi-purpose, standardized description. The latter is not treated in more detail here.

Demography & Population Studies - WWW Virtual Library	
Navigation	Guide Information
Search/Browse	http://demography.anu.edu.au/VirtualLibrary/
Main Page ▼	Keywords demographics, <u>demography</u> , population
Social Sciences & Social Issues ▼	Compiled by Diana Crow (Diana.Crow@anu.edu.au) Demography Program, RSSS, Australian National University
sociology ▼	G. Longmuir
population ▼	Rating Overall: Resource Description: 2 Resource Evaluation: 1 Guide Design: 4 Organization Schemes: 4 Guide Meta-information: 3 (Rated 02/2001)
Guide Information	Last Checked by Argus Clearinghouse June 20, 2000

Figure 1: Guide information in the ARGUS Clearinghouse

(from: Ohly, H. Peter: Gestaltungsprinzipien bei sozialwissenschaftlichen Wissensportalen im Internet. In: Lehner, Christoph; Ohly, H. Peter; Rahmstorf, Gerhard (Hrsg): Fortschritte in der



Wissensorganisation Vol. 7. Wissensorganisation und Edutainment: Wissen im Spannungsfeld von Gesellschaft, Gestaltung und Industrie. Proceedings 7th German ISKO conference, Berlin 2001. Ergon: Wurzburg, 2004, p. 234 - 246)

Procedures for quality assessment by a brokerage service are listed by the 'Collection Development Policy Framework' of Intute (2006) on the basis of the UK Resource Discovery Network (Jennings, 2006). Intute, an association comprised of over 70 organizations from education and science in Britain, is a free online service providing a database of hand-selected Web resources for education and research, covering Science and Technology, Arts and Humanities, Social Sciences and Health and Life Sciences. Regarding the item labelled 'Values' it states: "*Quality is our cornerstone; human expertise and value judgments are irreplaceable...* ", and for 'Aim' : "*The Intute collections are selective, and only resources that meet the Intute quality selection criteria are included ...* ."

The following methodical aspects are treated under 'Selection Criteria:' Scope (subject area, intended audience, acceptable sources of information, acceptable levels of difficulty, Internet resource types and categories, geographical, language), Selection Policy (resource discovery, selection and evaluation criteria), Collection Management and Maintenance (duplication, cross-disciplinary resources, granularity, optimum size of collections, reviewing records, link checking, deselecting records, contact with website / resource owners). The Internet resource types and categories are defined in more detail. For example, illegal activities, marketing activities and personal homepages are sorted out.

Of particular interest are the selection and evaluation criteria, which are divided into: subject knowledge, core criteria, intellectual content, objectivity, structure and form, system and maintenance, contributors, public suggestions. Here again, the intellectual content is given special significance: "*The intellectual content of a resource, the information contained within the website being evaluated, takes priority more over the form in which it is delivered or displayed.*" Aspects mentioned for consideration are: declared purpose and audience of the site, authority and reputation, accuracy of the information or knowledge presented, evidence of sources used and cited, originality in content or presentation, comprehensive coverage and depth of information, evidence of external citation.

All of these are relative prescriptions, which let the respective evaluator suppose a good knowledge of the brokerage database and the available Web sources. More explicit are the exclusion criteria, which demand deselecting:

- If the resource is no longer available.
- If the currency or reliability of the resource has lost its value.
- If another Internet site or resource offers more comprehensive coverage.



- If it is a duplicate record.

In summary, the following closer regulations of the main aspects: Context, content, use features and system features are given (see Table 3).

Table 3: Context Content Use features system features

Context	Content	Use features	System features
Provenance	Scope/coverage	Accessibility	Site integrity (availability, broken links etc.)
Authority	Accuracy	Navigability	System reliability
Uniqueness	Currency	Terms and conditions of use	Appropriate standards and technologies
Relationship to other resources	Substantiveness	Rights legitimacy	
Audience	Comprehensiveness	Design, layout, and aesthetics	
	Composition and organisation	User support and documentation	

The difficulty in determining something such as accuracy is shown by the following questions, which can be pursued with this intent: Where does the information come from? Is an individual or group responsible for the resource? Is an organization responsible for the information? Is the resource well known and/ or heavily used? Is there a feedback form to show that the site administrators are interested in what the users think? Does the URL provide clues to credibility? Who else links to the resource and are they reputable?

Ultimately, the recommendations add up to leaving the decision with the designated specialized editor as to which resources to exclude and how critically to give a detailed description for the selected sources by avoiding an explicit total valuation score. It is also the user who ultimately selects the description characteristics which allow appropriate information to be offered in response to his or her question (cf. http://www.intute.ac.uk/intute_cataloguing_guidelines_v4_1.doc). Accordingly, also for the various professional services different evaluation guidelines are allowed, thus for Social Sciences as well as for Health and Life Science. Whereby the latter gives more elaborated instructions, e.g. under the point 'Reasons for not including resources' (Intute, 2006-2008).



3. THE GESIS SOCIOGUIDE

3.1. Selection Policy

The aim of the subject gateway GESIS SocioGuide is to offer orientation within social science infrastructure and information from all social science disciplines as well as on special topics. For that purpose, science-relevant resources in a categorized form have been made accessible in GESIS SocioGuide, gathered from the pool of available web-based resources and other media; which are not subject to discussion here. The subject gateway GESIS SocioGuide uses the workflow-system based on DBClear (Hellweg et al., 2002).

The SocioGuide catalogue includes descriptions of the following types of Internet resources relevant to the subject coverage and primary audience: institutions, journals, networks, collections, and events. The GESIS SocioGuide currently lists 15,000 records (as of November 2008) and provides access to detailed and regularly updated descriptions of resources such as social science journals or (German and Eastern European) social science research institutions. The extent of coverage depends on information types, languages and geographical regions. In stocks such as networks, collections and events, the collection profile was limited because of the numerous information offers. The metadata structures have been adjusted to the specifics of the referred resources.

The GESIS SocioGuide is to be further developed by the addition of individualization components slated to provide user-oriented current services.

Social scientists are the primary target group of the GESIS SocioGuide. This, of course, does not preclude other academics, politicians, decision makers and many other professionals from using social science data and benefitting from this subject gateway.

GESIS SocioGuide (priority) applies the "three click rule," whereby the formal topic of a resource should be obvious within three clicks (based on Bargheer, 2002). The following aspects are supported:

- Origin (providing institution, country/ regional scope),
- Content (information type),
- Degree of complexity (conditionally operational).

To sum up, the work in the subject gateway GESIS SocioGuide is characterized by the following components (cf. Rösch/ Weisbrod, 2004):

- Intellectual selection of document resources by editors in charge according to defined quality criteria
- Use of standardized metadata for the description of the selected sources, whose field structure far exceeds the metadata structures of Dublin Core (Dublin Core Metadata Element Set, see <http://dublincore.org/documents/dces/>)
- Regular check-up of (selected) links regarding their relevance, validity, and longevity



- Re-examination of information resources' content as to document descriptions
- Subject indexing by controlled, vocabulary respective specialized classification
- Co-operative working in the database
- Integration of document types in sowiport (<http://www.sowiport.de>) as a subproject within the social science portal sowiport.

Data and metadata are stored in a database. The subject gateway has been designed to meet interoperability features which permit the integration/ representation of the data sets in other services, such as sowiport (<http://www.sowiport.de>), vascoda (<http://www.vascoda.de/>) and Knowledge Base Social Sciences Eastern Europe (<http://www.cce-socialscience.net/>).

3.2. Semi-Automatic Quality Control by Workflow – Components

The workflow has been implemented as the coordinated management of interdependent steps in the working process in the environment, or under the conditions of a work-sharing process. Each process step is managed by different editors. This helps to limit the routine activities and to standardize the procedure.

Additionally, workflow is partially a system-specific infrastructural tool whose supporting function is very important. Automation and control are of primary importance for the optimization of the work process.

"To facilitate the creation and continuous maintenance of a data collection that exceeds a certain size, tools for automation of tasks are required, as well as support for the distribution of work among several cooperating editors" (Hellweg et al., 2002). But relying only on automatic methods does not ensure the high quality standard needed. *"The software system DBClear provides a number of modules to automate recurring tasks and a workflow system to route information between the people involved in a clearinghouse"* (Hellweg et al., 2002).

In order to meet the diverse requirements (high quality and a multitude of resources) we are choosing a strategy of semi-automatic quality control. For this purpose, records are regularly reviewed alongside the website it describes. Records for review are identified automatically. This involves a regular, weekly-automated check of all URLs contained in the catalogue records. In the following we will consider some essential, characteristic features of the GESIS SocioGuide.

Practical components of the quality assured workflow are:

Resource discovery: Comprehensive and high quality development of scientifically relevant sources of information may be realized under the current international debate only with intellectual assessment and development. Search engines and automated procedures can be used for specific requests. Nevertheless, there are no alternatives to the intellectual search if the exploration and selection of relevant information sources of a complete subject field is required.



The information for the subject gateway SocioGuide is obtained in several ways:

- Online discovery tools such as trusted Internet gateways
- Following links within trusted sites,
- Reading lists, email lists and discussion groups,
- Expert interviews,
- Literature and field analysis, or by
- Resources suggested by users, partner organisations, and staff.

• **Editorial qualitative validation**

New entries as well as existing entries are regularly validated by a professional editor.

The task of the check-up/control of the sources on validity and how current it is, of special significance to ensure the quality the GESIS SocioGuide within the workflow processes. While creating a new dataset a duplicate check is carried out. Then the editor in charge makes a cost-benefit oriented indexing or updating of the document.

All Internet resources should offer direct access to the described content. The sources in the subject gateway GESIS SocioGuide are mainly subject to the following criteria/tests:

A) Scope and Content Criteria

- Relevance as to the collection profile, i.e. document type, scope, language, geographical area
- Availability of an English-language homepage (otherwise national language homepages with extensive information offers)
- Longevity and validity of the information

Provided that the information offered for the most part is outdated or contains "dead links," the resource is then given careful consideration and optionally not included.

B) Form Criteria

- The possibility to contact the owner of the source
- Access to the source must be provided with the usual hardware and software, the sources that require downloading unusual plug-ins or certain browser versions are thoroughly checked.
- Sources difficult to access or under construction are not recorded.

Besides the formal and content-oriented characteristics further criteria (process criteria and criteria of collection management policy) are applied for a detailed check of the sources (cf. Quality Selection Criteria for Information Gateways, <http://www.sosig.ac.uk/desire/qclont.html>).

During the process of the cataloguing of new titles/ records the additional information can be extracted from the protocol information (HTTP header) a web server sends with each page or from the URL, which can contain country codes or institution names (see Hellweg et al., 2002).



• Editorial automated validation

Depending on the amount of data, the editors need support in automated management and viewing. The entries are checked by automated routines to ensure integrity during storage. Several aspects of gateway maintenance have been automated. The most obvious are regular checks to see whether a website of a resource can still be reached (link checker) or has been modified since the last time it was checked by an editor. The documented links are examined daily.

Furthermore, the system verifies that all required fields are filled. A to-do-list contains the tasks assigned to a person or a group. To coordinate the tasks, the system provides information on the state of a resource along with a work list for each editor.

- The automatic/ manual workflow is initiated by messages both from the system and the editor in charge.
- Different additional areas which cause an automatic re-submission of documents can be created in the Worklist. On the one hand, a field like "last update" of a dataset (comparison with the corresponding source) or the event date can serve as a criterion. The message of the re-submission is addressed to the owner or the group in charge, whose corrector completes the existing record.

◆ Suggestions by users

Users may recommend web entries of their choice. They are later processed by the professional staff of GESIS. Figure 2 shows a simple workflow sequence, in which a user's suggestion is categorized and assigned to a stock, based on the information supplied. The system generates a message to the defined person in charge. By using this initial information, an editor (or a group of editors) is selected for entering the formal description of the resource, like country of origin, language or resource type. The editor (or a group of editors) will be informed by a message from the person in charge. If this selected editor decides that the automatic assignment to the stock was incorrect, he is free to reassign the resource to some other stock. Once the formal information is entered, the resource is ready for the content analysis (e.g. writing an abstract) and indexing. The following translation step has to be performed by an editor with the required language competence. The final publication of the resource on the Web is performed by the editor responsible for the consistency of the collection. (Hellweg et al., 2002).

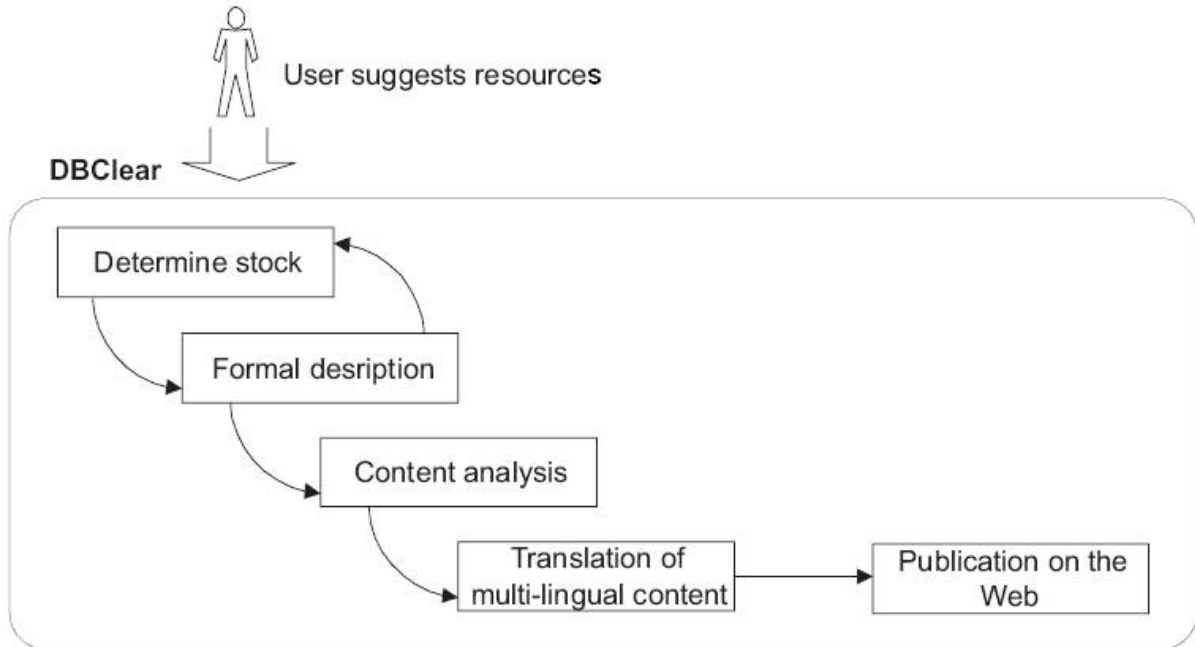


Figure 2: Example workflow sequence in DBClear

The workflow can be organised flexibly according to the possibilities of the institution, e.g. how many people have to be entrusted with the indexing of resources and which tasks are assigned to these people.

• Decentralized workflow

The processing of data records is also possible by job sharing between several facilities. At present the decentralized workflow has been implemented as several processing steps within an institution.

The widespread collection maintenance currently profits from the collaboration of different departments in GESIS as well as from the network collaboration with other institutions, such as *Wissenschaftszentrum Berlin für Sozialforschung* (Social Science Research Centre Berlin, <http://www.wzb.eu/>) and *Universitäts- und Stadtbibliothek Köln* (University and City Library Cologne, <http://www.ub.uni-koeln.de/>).

This collaboration is mainly based on compliance with defined principles and standards as well as on the willingness for active participation in the collection of resources.

The future cooperative interaction in the subject gateway GESIS SocioGuide will adhere to the principle of confident cooperation and decentralize the responsibility for quality assurance. Coordination in the following aspects is therefore of particular importance: agreement on the further development of common standards; clear definition of responsibilities, simple application of indexing tools, personal support and guidance, control.



4. CONCLUSION

In accordance with the state of the art, quality control of Internet sites which are included in subject gateways is essential. It must refer to exclusion criteria, description elements, workflow and technical interoperability. Automatic checks are tools for professional editors as well as for users. They cannot, however, be applied exclusively for quality evaluation. Therefore, in addition to regular control routines, informative, standardized descriptions are also necessary, which are elaborated according to Guidelines. The GESIS-SocioGuide correspondingly represents a semi-automatic procedure set to broker high-quality specialized Web sources, focusing on storage, maintenance and transfer.

5. REFERENCES (all URLs refer to status 1.12.2008)

- BARGHEER, Margo (2002): *Qualitätskriterien und Evaluierungswege für wissenschaftliche Internetressourcen*. Report zum DFG-Projekt. „Datenbankbasierte Clearinghouses im Kontext digitaler Bibliotheken. Niedersächsische Staats- und Universitätsbibliothek Göttingen. Available at: <<http://webdoc.sub.gwdg.de/ebook/aw/2003/bargheer/v10.pdf>>
- DESIRE (1996): Development of a European Service for Information on Research and Education, RE 1004, D3.2 2, Specification for resource description methods Part 2: Selection Criteria for Quality Controlled Information Gateways, Project Deliverable, Vers.1.1. Available at: <<http://www.ukoln.ac.uk/metadata/desire/quality/quality.rtf>>
- EL-MENOUAR, Yasemin (2002): Was erwarten Nutzerinnen und Nutzer vom Internet-Angebot medizinischer Bibliotheken? Ergebnisse einer internetbasierten Umfrage. In: *medizin-bibliothek-information*, Vol. 2, May, pp. 24-32. Available at: <http://www.meduniwien.ac.at/agmb/mbi/2002_2/24-32elmenouar.pdf>
- ENDERLE, Wilfried (2001). Der Historiker, die Spreu und der Weizen. Zur Qualität und Evaluierung geschichtswissenschaftlicher Internetressourcen. *Geschichte und Informatik/Historie et Informatique*, 12, 49-63.
- EPPLER, Martin J. (2006): *Managing information quality*, New York: Springer, (2.). ISBN: 3-540-23571-3
- EPPLER, Martin J., WITTIG, Dörte (2000): Conceptualizing Information Quality: A Review of Information Quality Frameworks from the Last Ten Years. In: *Proceedings of the 2000 Conference on Information Quality*. Available at:



<<http://mitiq.mit.edu/iciq/Documents/IQ%20Conference%202000/Papers/ConceptIQaReviewofIQFramework.pdf>>

- FOLLET REPORT (1993), Joint Funding Council's Libraries Review Group: Report, External Relations Department. HEFCE, Bristol, December 1993. Available at: <<http://www.ukoln.ac.uk/services/papers/follett/report/>>
- HELLWEG, Heiko, HERMES, Bernd, STEMPFHUBER, Maximilian, ENDERLE, Wilfried, FISCHER, Thomas (2002): DBClear: A Generic System for Clearinghouses. In: Adamczak, Wolfgang; Nase, Annemarie (eds.): *Gaining Insight from Research Information: Proceedings of the 6th International Conference on Current Research Information Systems*, University of Kassel, August 29-31, 2002. Kassel: Kassel University Press. pp. 131-139. Available at: <<http://www.uni-kassel.de/CRIS2002/files/pdf/Hellweg.pdf>>
- HOB OHM, Hans-Christoph (1998): *Qualität und Nutzerorientierung bei bibliothekarischen Internetprojekten – marketingstrategische Überlegungen zu den neuen Informationsdienstleistungen*. Available at: <<http://www.fh-potsdam.de/~hobohm/inetqual.pdf>>
- HOB OHM, Hans-Christoph (2003): *Qualitätsempfehlungen im Internet*. Available at: <<http://dspace.hrz.uni-dortmund.de:8080/bitstream/2003/2197/1/hobohm.pdf>>
- HUANG, V, LEE, Yang W. ,WANG, Richard Y. (1999): *Quality information and knowledge* Prentice Hall: Englewood Cliffs, NJ.
- INTUTE (2006), *Collection Development Framework and Policy*. Version 1.4 Approved by the Intute Policy and Strategy Forum 12/10/06. Available at: <http://www.intute.ac.uk/Intute_cdfp.doc>
- INTUTE (2006-2008), *Health and Life Sciences Evaluation Guidelines*. Available at: <http://www.intute.ac.uk/healthandlifesciences/IntuteHLS_Evaluation_Guidelines.doc>
- JENNINGS, Simon (2002): *RDN Collections Development Framework*. Version 1.2. July, 2002. Available at: <<http://www.webarchive.org.uk/pan/14883/20060912/www.rdn.ac.uk/publications/collections/cdframework3.doc>>
- KAHN, Beverly K., STRONG, Diane M. (1998): *Product and Service Performance Model for Information Quality: An Update*. In: *IQ 1998*, pp. 102-115.
- KOCH, Traugott (2000): *Quality-controlled subject gateways: definitions, typology, empirical overview*, In: *Online Information Review*, Vol. 24, No. 1, pp. 24-34



- KUHLEN, Rainer (1995): *Informationsmarkt: Chancen und Risiken der Kommerzialisierung von Wissen*. Konstanz: Universitätsverlag Konstanz. Available at: <<http://www.inf-wiss.uni-konstanz.de/FG/IV/imk/imk.html>>
- MANDL, Thomas (2005): The quest for the best pages on the web. In: *Information Services & Use*. Vol. 25 (2), pp. 69-76.
- NOHR, Holger (2001): Management der Informationsqualität. In: *Working Papers Knowledge Management*, Nr. 3/2001, ISSN 1616-5349. Available at: <<http://www.iuk.hdm-stuttgart.de/nohr/KM/KmAP/Informationsqualit%E4t.pdf>>
- OHLY, H. Peter (1998): Structuring of Social Science Information Resources in the Internet: Knowledge-Organization Aspects in Building a Clearinghouse International Conference: 25-27 March 1998, Bristol, UK. Conference Paper IRISS '98. Available at: <<http://www.intute.ac.uk/socialsciences/archive/iriss/papers/paper51.htm>>
- OHLY, H. Peter (2004): The Organization of Internet Links in a Social Science Clearinghouse. In: McIlwaine, Ia C. (eds.): *Knowledge Organization and the Global Information Society*; Proceedings of the Eighth International ISKO Conference, 13-16 July 2004, London, UK. Würzburg: Ergon Verl. (Advances in Knowledge Organization; Vol. 9). pp. 121-126.
- OHLY, H. Peter (2008): Quality and Information – an introduction. Talk at the RC33 Seventh International Conference on Social Science Methodology, Naples, Italy.
- RITTBERGER, Marc. (2004): Vertrauen und Qualität in Informationsdienste. Wo finde ich Vertrauen im Information Quality Framework? In: Hammwöhner, R., Rittberger, M. und Semar, W. (eds.): *Wissen in Aktion. Der Primat der Pragmatik als Motto der Konstanzer Informationswissenschaft. Festschrift für Rainer Kuhlen*. Konstanz: UVK Verlagsgesellschaft. pp. 153-165. Available at: <http://www.inf-wiss.uni-konstanz.de/infwiss/download/festschrift/cc-festschrift_RK-art12.pdf>
- RÜTER, Martina (2008): Qualität von Webseiten Available at: <http://tools-internetservices.suite101.de/article.cfm/qualitaet_von_webseiten>
- RÖSCH, Hermann, WEISBROD, Dirk (2004): „Linklisten, Subject Gateways, Virtuelle Fachbibliotheken, Bibliotheks- und Wissenschaftsportale: Typologischer Überblick und Definitionsvorschlag“. In: *BIT-online*. 7, 3, S. 177-188. Available at: <<http://www.b-i-t-online.de/archiv/2004-03/fach1.htm>>
- RÜTER, Martina (2008): *Qualität von Webseiten*. Available at: <http://tools-internetservices.suite101.de/article.cfm/qualitaet_von_webseiten>



- STRONG, Diane M., LEE, Yang W., WANG, Richard Y. (1997): *Data quality in context*, Communications of the ACM, v. 40, n. 5, p.103-110. Available at: http://www.ir.iit.edu/~dagr/DataMiningCourse/Research_Papers/p103-strong.pdf
- WANG, R. Y. and STRONG, D. M. (1996): Beyond accuracy: *What data quality means to data consumers*. J. Manage. Info. Syst. 12, 4, pp. 5–34
- WERSIG, Gernot (1973): *Informationssoziologie: Hinweise zu einem informationswissenschaftlichen Teilbereich*. Frankfurt (am Main): Athenäum-Fischer-Taschenbuch-Verlag.
- WIETHAUS, Jörg (2001): *Entwicklung und Nutzung eines Information Quality Rating-Tools am Beispiel von Content Management-Systemen*, Master Thesis Univ. Konstanz. Available at: <http://www.ub.uni-konstanz.de/kops/volltexte/2002/758/pdf/Wiethaus.pdf>
- WISSENSCHAFTSRAT (2001): *Empfehlungen zur digitalen Informationsversorgung durch Hochschulbibliotheken*, July 13. Available at: <http://www.wissenschaftsrat.de/texte/4935-01.pdf>