


First Monday, Volume 12 Number 12 - 3 December 2007

 First Monday

A model for ontology quality evaluation

by Besiki Stvilia

 Abstract

Ontologies are important knowledge representation and sharing tools in the workflow of biology research. The high cost of creating and maintaining ontologies encourages their sharing and reuse, and an increasing number of ontologies have been made available from different sources, with different models of curation. To enable effective selection, reuse, integration, and maintenance of these ontologies, however, one needs to have a systematic method of evaluating and connecting their quality to the context of an intended use. Based on an analysis of the activity system and Web server logs of the Morphbank biodiversity research data repository (<http://www.morphbank.net/>), a model was developed to evaluate ontology quality. The model connects the types of quality problems with the types of research activities and suggests relevant metrics. The paper also describes the structure of some of the research activities and the types and patterns of end-user searches in Morphbank.

Contents

[Introduction](#)

[Related research](#)

[Research setup and method](#)

[Analysis](#)

[Constructing a quality evaluation model](#)

[Conclusion](#)

Introduction

In this paper, we develop a model to evaluate the quality of a biodiversity ontology. The model is an operationalization of the information quality (IQ) assessment framework proposed earlier (Stvilia, 2006), which combines conceptual and empirical approaches to identify an IQ problem structure and the requirements for an information object and grounding IQ metrics.

With advances in the Internet and the Web, as both information-knowledge creation workflow and services move increasingly to the Web and as they become decentralized and are distributed, the process of ontology construction and maintenance is distributed as well. The Web may also encourage the exposure and aggregation of local metadata and ontologies that might not have been created with sharing in mind. An increasing number of projects are collecting research data encoded in different metadata schemas (e.g., Jones, *et al.*, 2001) that provide integrated workbenches for scientific work, including tools for workflow management and analysis (e.g., Ludäscher, *et al.*, 2006), or that build mashups of different ontologies and reference sources to provide task-specific knowledge support (e.g., Parr, *et al.*, 2006). Web portals have been established that collect and provide access to research data, metadata, ontologies, literature, and software tools (e.g., National Biological Information Infrastructure at <http://www.nbio.gov/>, National Center for Biotechnology Information at <http://www.ncbi.nlm.nih.gov/>).

These repositories and portals contain an immense wealth of knowledge and data, but effective integration and reuse of this knowledge and these data have proved to be difficult (e.g., Lagoze, *et al.*, 2006). Their interoperability and reusability can be hampered not only by

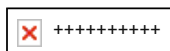
differences in the cultures, reference sources, and pragmatics of the respective communities and organizations, but also by the lack of robust models and mechanisms for evaluating the quality of these sources effectively and inexpensively.

Traditional libraries have used a centralized approach with trained catalogers and indexers to create, maintain, and use controlled vocabularies and classification systems or taxonomies, such as LCSH, MeSH, Dewey, or LCC. Although this approach has been successful for libraries, it may not be affordable or applicable to the less formalized information workflow and more transient information objects and knowledge sources used in scientific laboratories. This does not mean, however, that the quality of controlled scientific vocabularies and ontologies is not important to the quality of scientific work. The quality of an ontology may have a direct impact on the quality of an outcome of the processes in which it is used (see Köhler, *et al.*, 2006).

Having a systematic and standardized method of evaluating the quality of research data and reasoning about research conclusions can benefit not only scientists, but also government officials and policy makers, who may aggregate and use this information in making their own decisions or in justifying previously made decisions (Gasser, 2003).

Several digital library projects (*e.g.*, OAI — Object Reuse and Exchange (ORE), DLF — Aquifer) developing interoperability architectures to support cross-repository workflow of scholarly digital objects (Van de Sompel, *et al.*, 2006). The architecture proposed by the Pathways project (<http://www.infosci.cornell.edu/pathways/>) includes a shared data model to represent digital object and three core repositories for formats, services and semantics. Although currently the architecture does not include quality evaluation mechanisms, if implemented, the data model and repositories can be used as reference sources for quality evaluation models.

The need exists for a shared standardized vocabulary, for typologies of quality problems and activities, and for a taxonomy of dimensions — an IQ ontology. In addition, the need exists to develop information type-specific models of quality assessment, with type-specific metrics and baseline values. In an earlier study, we developed a general framework — an IQ knowledge base — that can be reused to develop information type- and context-specific IQ measurement models. In this work, we conducted a content analysis of specimen image annotations and Web server search logs of the Morphbank biodiversity data repository and laboratory to identify the structures and types of some of its information activities and investigate the need for quality in its content and tools. The findings of the analysis, combined with insights gained from a literature analysis, were then used to operationalize the framework and develop a model to evaluate a biodiversity ontology.



Related research

Approaches used for evaluating ontologies can be grouped into four general categories. The first approach uses information about the class graph of an ontology to measure the complexity of the ontology. The second approach analyzes the lexical or linguistic structure of the content of an ontology to evaluate its quality characteristics, such as clarity, interpretability, and redundancy. The third approach exploits an outside evaluation mechanism, such as Google's PageRank, or uses human reviewers and the positions of reviewers in a social or trust network to assess the quality of an ontology indirectly. The fourth approach uses a combination of the three approaches.

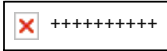
Burton-Jones, *et al.* (2005) proposed a set of nine ontology quality metrics grounded in the lexical or linguistic structure of an ontology and its popularity in an activity system: lawfulness, richness, interpretability, consistency, clarity, comprehensiveness, accuracy, relevance, and authority.

In a similar manner, Köhler, *et al.* (2006) proposed the metrics of circularity and intelligibility to evaluate the quality of the Gene Ontology (<http://www.geneontology.org>). The metrics use the linguistic structure of the ontology entries and the WordNet ontology (<http://wordnet.princeton.edu/>) as a reference source to compute quality scores.

Orme, *et al.* (2007) compared automatic measurements and human evaluations of tourist ontologies and found that the software complexity metrics (*e.g.*, number of properties, average properties per class, average fan-out per class) could be successful in automatically evaluating the complexity and cohesiveness of an ontology. Lewen, *et al.* (2006) used a

recommender system approach to evaluate ontologies. They developed a system that rates ontologies based on user reviews and a reviewer trust relationship network.

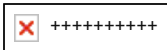
Although a significant amount of research has been conducted on ontology development (e.g., Corcho, *et al.*, 2003) and ontology evaluation, there is still a need to develop a systematic and structured method for assessing ontology quality. The previous studies have substantially advanced research on ontology quality evaluation by providing empirical data on quality variance sources and proposing quality metrics. However, with the exception of Burton-Jones and colleagues (2005), little work has been done in developing systematic models for ontology quality evaluation. Having a systematic and consistent method of evaluating IQ is essential for quality-based cross-comparison and decision making, and for the ease of measurement model construction and reuse.



Research setup and method

The theoretical framework used for this research consisted of activity theory (Engeström, 1990; Leont'ev, 1978) and an IQ assessment framework developed earlier (Stvilia, 2006). The theoretical framework helped us to conceptualize the activities of creation and use of a biodiversity ontology: their structure, the need for quality, and the types of quality problems to which the ontology could be prone. The conceptualization of the ontology's activity system and the suggested IQ problem structure were then used to guide an empirical analysis of specimen image annotations and Web server logs in Morphbank to produce a model for evaluating the quality of this biodiversity ontology.

In the empirical analysis, we looked at 378 distinct annotations and comments on specimen images in Morphbank to identify the requirements of the community and the need for tools and quality. In addition, to identify search types and vocabulary patterns, we examined the Web server request logs from April and May of 2007. The log files were preprocessed before analysis to reduce the footprints of spam and Morphbank maintenance activities. All requests made by the network domain of the Morphbank project and by search engine bots and known spammers was removed, which reduced the number of Web server requests available for analysis almost 10-fold — from more than 500,000 to less than 49,000.



Analysis

In this study we used the information theoretical framework (Engeström, 1990; Leont'ev, 1978) for IQ assessment, which was developed previously (Stvilia, 2006) to construct a conceptual model of ontology quality. The conceptual model was then used to guide an analysis of empirical data to identify instances of quality problem types and specific needs for vocabulary and content quality in Morphbank.

Morphbank (<http://www.morphbank.net/>) is a quickly growing research collaboratory and repository of biological images. The data include specimen taxonomy information, images, morphological character matrices, and annotations. The Morphbank community includes 170 registered researchers from around the world. The repository contains more than 70,000 specimen images submitted by different content providers, representing approximately 5,000 unique taxa.

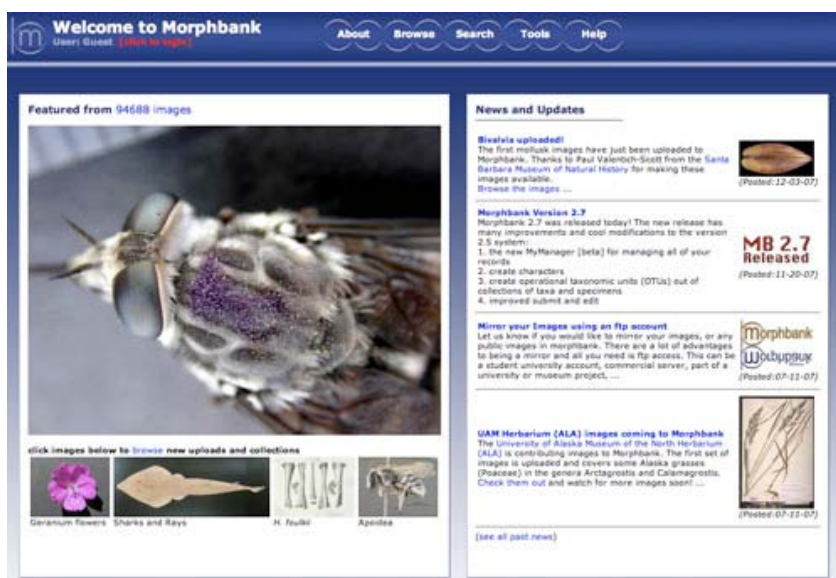


Figure 1: The opening page of Morphbank (<http://www.morphbank.net/>), accessed 6 December 2007.

The Morphbank system is intended not only as an archival or repository tool, but also as a workbench for comparative anatomy, taxonomy, and morphological phylogenetic research. It aims to provide an integrated environment by combining research data, metadata, literature, and ontologies created by its content providers and partners (e.g., *Hymenoptera* ontology at <http://ceb.csit.fsu.edu/ronquistlab/ontology/>), as well those created by other biology or biodiversity communities and made publicly available on the Web (e.g., the BioPortal of the National Center for Biomedical Ontology at http://www.bioontology.org/ncbo/faces/pages/ontology_list.xhtml).

Activity theoretic conceptualization of quality

The most widely used definition of quality is “fitness for use” (Juran, 1992). Quality is a multidimensional and contextual concept (Strong, *et al.*, 1997). According to activity theory, context can be viewed as an interplay between general cultural and community structures (language, norms, conventions, social networks, and relationships) and the structure of a particular activity or activities (actions, goals, needs or requirements, tools, roles, rules, strategies, etc.; see [Figure 2](#)). Hence, to measure the quality of an ontology, or any information object for that matter, one needs to (1) understand the general cultural and community context of the ontology; (2) understand the context of specific activities of ontology creation, use, and maintenance; and, (3) define mechanisms that can capture variances in the ontology attributes, their underlying entities, and their relevant contextual (cultural, activity) features and relations, and transform them into quality measures or metrics.

Information quality measurements can be grouped into three categories or groups: *intrinsic*, *relational*, and *reputational* (Stvilia, 2006). Intrinsic quality measures the internal characteristics of the information itself in relation to some general reference standard in a given culture, such as WordNet, for instance. Relational or contextual quality measures relationships between the information and some aspects of its usage context and its reference sources. Indeed, some of the IQ dimensions (e.g., completeness) can be evaluated only in relation to the needs of a specific activity or action. Finally, reputational quality is the position of information in a cultural or activity structure (e.g., a trust network), often determined by its origin and record of mediation.

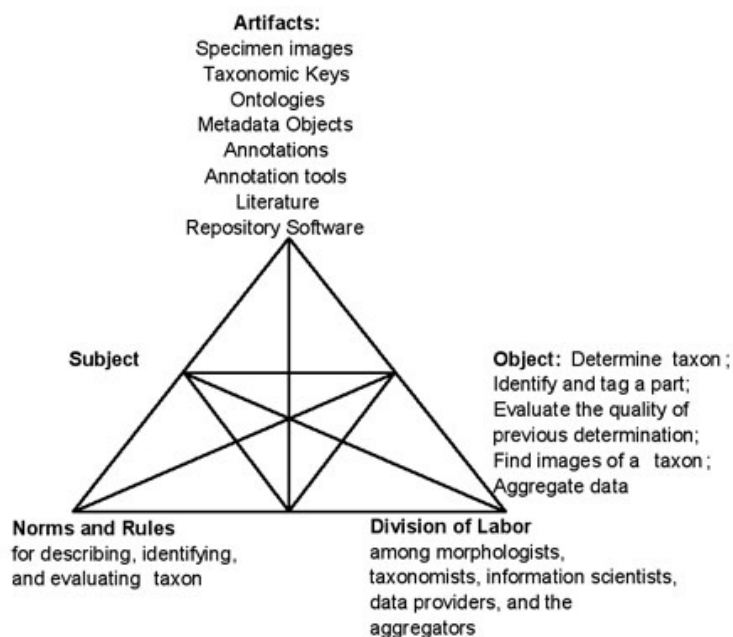


Figure 2: Morphbank's activity system.

A quality problem arises when the *existing* quality level is lower than the *required* quality level for a given activity. Quality problems may arise in any of the three groups: intrinsic, relational, or reputational. Furthermore, quality levels in all three groups may change with changes in the underlying entities or in the overall context of information evaluation or use, leading to dynamic IQ problems. Thus, an effective IQ model needs to measure variances in the levels of intrinsic, relational, and reputational quality, as well as to evaluate their dynamics in time and space (dynamic IQ problems).

Some activities may depend on how well an information object or an information repository as a whole represents some external situation or process. Other activities may remove information from the context in which it was produced. For example, an activity may aggregate raw information from a variety of original sources and integrate it into a focused collection supporting a specific task. Other activities may depend on the stability of the information properties or on the entities and conditions it represents (*e.g.*, referencing). Finally, activities may depend on the state of an information object at a particular point in time and space (*e.g.*, archiving and restoring). Therefore, these activities may depend on the availability and quality of the provenance and manipulation records of an information object.

In general, an ontology is defined as the specification of conceptualizations of a domain — definitions of classes, relations, and functions of a vocabulary for a shared domain (Gruber, 1993). Hence, an IQ measurement model for an ontology needs to evaluate the quality of the mappings of domain concepts into the classes, relations, and vocabulary of the ontology. That is, the model needs to measure how completely, consistently, or accurately the ontology represents the domain concepts in relation to the general cultural context and the context of a particular activity system. In addition, the model needs to evaluate the probabilities of change in the ontology and its context, and the effects of change on the quality evaluations of the ontology. The model should be able to evaluate how stable the ontology is. Changing taxonomic definitions and relations may invalidate outcomes of the activities that used or referenced that data. Note that there could be a trade-off between the goals of stability and representational quality. To maintain the representational quality of an ontology, the ontology needs to be updated and aligned regularly with changes in its underlying entities and context. The updates, however, may have a negative impact on the activities and artifacts that depend on the stability of the ontology.

The effects on quality of a community or cultural context change can be more significant than the effects of an activity context change. Indeed, when moving an ontology from one community or cultural context to another, all baseline models and reference sources one uses to evaluate the quality of an ontology may change. A simple example would be using a scientific ontology of plant information to expose young children to biodiversity information at a field museum. This may require changing not only the baseline models for the quality dimensions, such as completeness or complexity, but also changing reference sources, such as

the vocabulary or language. Likewise, a plant ontology used by morphologists may not meet the needs of geneticists. Changes in the activity context, on the other hand, may affect only baseline models. For example, a taxon identification or determination activity may require a higher degree of completeness of description than the activity of annotating or tagging specific parts of a specimen.

Decision-making activities are usually provenance dependent. When making a decision, it is often important for a decision-maker to have access to the provenance record of information — who did what, when, and in what context — to evaluate the trustworthiness or reliability of the information. An ontology produced and curated by qualified specialists at a reputable institution is expected to be of high quality. In addition, there can be a need to restore the ontology to a specific state in time, to recover from an unqualified edit, for instance.

Finally, quality metrics can be attribute or process based. The quality of an ontology can be measured either directly, by measuring the variance in its attributes (e.g., number of ontology classes), or indirectly, by evaluating the quality of its construction and maintenance processes. Some of the metrics for evaluating process quality could be the qualifications of its editorial body or the presence of quality assurance mechanisms.

Information activities in Morphbank

A biodiversity taxonomy and systematics research process may consist of the following tasks: *collecting*, *naming*, *describing*, and *classifying* specimens (Hodkinson and Parnell, 2007). One may also need to evaluate or comment on existing taxonomic information, and *share* or *exchange* information with other parties. Each of these tasks can be accompanied and supported by a set of distinct information tasks and tools, including ontologies. Furthermore, information, knowledge, and metadata generated throughout the research process need to be *captured*, *organized*, *documented*, *preserved*, and made easily *accessible* and *reusable*. A research process may also include *searching*, *synthesizing*, and *integrating* existing literature and data, as well as *presenting* and *disseminating* research results and findings (Crawford, *et al.*, 1996). Each of these activities may use an ontology or ontologies.

There are two distinct methods of knowledge system construction (Bailey, 1986). The classic, top-down or deductive, approach first develops theoretical or hypothetical concepts or relations, which are then mapped onto their empirical examples. The second, bottom-up or inductive, approach first identifies empirical clusters and associations and then assigns conceptual labels to them. Bioinformatics researchers have been using both the top-down and the bottom-up approaches, as well as various combinations of the two, to construct ontologies and other knowledge systems (see Leroy and Chen, 2005, for a review). Similarly, quality requirements can be identified top down through conceptual analysis of an organization's activity system, or bottom up by analyzing exemplary data collections or archived process data (Stvilia, 2006).

Gruber (1995) proposed the following criteria for ontology construction: (1) Clarity: ontology definitions should be objective and independent of the social and computational context; (2) coherence: inferences drawn from the ontology must be consistent with its definitions and axioms; (3) extendability: the ontology should be designed to anticipate the uses of the shared vocabulary; (4) minimal encoding bias: conceptualization of the ontology should be specified at the knowledge level and be independent of symbol-level encoding; and, (5) minimal ontological commitment: an ontology should require the minimal ontological commitment sufficient to support the intended knowledge-sharing activities.

According to Chapman (2005), biology taxonomic data may contain the following: (1) name (scientific name, common name, hierarchy, rank); (2) nomenclatural status (synonym, accepted, typification); (3) reference (author, place and date of publication); (4) determination (by whom and when the record was identified); and, (5) quality fields (accuracy of determination, qualifiers). In addition, Chapman identified eight types of related information activities and tasks: (1) data capture and recording at the time of gathering; (2) data manipulation prior to digitization (label preparation, copying of data to a ledger, etc.); (3) identification of the collection (specimen, observation) and its recording; (4) digitization of the data; (5) documentation of the data (capturing and recording the metadata); (6) data storage and archiving; (7) data presentation and dissemination (paper and electronic publications, Web-enabled databases, etc.); and, (8) using the data (analysis and manipulation).

The data stored in the morphbank database are images of specimens that have already been collected, and often specimens that have been described as well. Hence, in examining image annotations, we identified only the following activities that Morphbank users might be engaged in: *determining* the taxon of a specimen, *tagging parts* of the taxon, *identifying and tagging anomalies* or unusual characteristics, and *evaluating the quality* of a particular determination or of taxonomy data.

The majority of annotations referred to determining the taxon of a specimen:

This is the genus *Ceraphron*.

In some cases, an object of the determination or identification activity was an individual part of the specimen. Members might mark or tag important morphological characteristics of a taxon that could be used as keys in a taxon determination process or in similarity searches:

Labelled feature is a discocellular spot.

Leaves are flabelliform (fan shaped) and not auriculate (lobed at base). This is last choice in *Flora of Australia* key.

This wing cell is diagnostic for the genus *Evaniscus*.

Morphbank members might also reevaluate or assess the quality or certainty of a prior determination:

This is not *Evania albofacialis*; it is the North American sp. *Evaniella semaeoda* (not from Costa Rica).

I concur with this determination.

Very likely *J. Americana*, but lacking inflorescences it is not possible to be 100% certain.

Finally, members might study anomalies in a specimen — features that would be unusual to its taxon. The presence of anomalies could be an indicator of a hybridization process for the taxon. Alternatively, their presence could be the result of a taxonomy error:

The *Justicias* of Florida may be over-split. The characters that ostensibly separate species are extremely subtle and cannot be seen in most of the images presented here. The differences also vary among floras, indicating different interpretations of taxa and suggesting that work is needed.

When making a taxon determination or tagging a part of a specimen, members would rely on various reference tools, such as flora catalogs or taxonomic keys. In addition, citing reference sources could serve as a means of signaling or indicating the quality of an identification process:

Involucral bracts of *B. serratulooides* (= *Dryandra serratulooides*) described as shining brown in *Flora of Australia* (p. 290).

According to our conceptual model of the Morphbank information processes (see [Figure 2](#)), the quality of the outcome of a determination activity could be affected by the quality of representation of a specimen, the quality of available tools, and the capability or expertise of the taxonomist. The annotations contained numerous references to uncertainties caused by poor quality or by an incomplete or inappropriate set of images for a specimen:

Can't be determined to variety given the images available, image of close-up of achene is essential to this det.

Only var. *muehlenbergii* is supposed to show up in FL, but without seeing the abaxial face of the perigynium, I can't be sure.

In addition, the activity could fail if the specimen itself was a poor representative of the taxon, either too old or undeveloped:

Identification may be correct, but specimen is depauperate.

Probably; overmature and difficult to see some features.

Outdated or incomplete reference keys and catalogs could lead to a failed or inaccurate identification:

These specimens are *Carex aureolensis* Steudel, not *C. frankii*. *Carex aureolensis* is a validly published name, and one in use in *FNA*, vol. 23, but does not appear in the ITIS list.

Finally, a determination made by an expert after consulting an authoritative source could be perceived to be of higher quality (certainty) than a determination made by a student using the same reference source.

Vocabulary needs

An analysis of Web server logs helped us to identify some of the characteristics and patterns of vocabulary used in Morphbank searches. The findings of this analysis could inform the structure and content of a Morphbank ontology or ontologies. Alternatively, they could help to identify some of the quality requirements and inform the design of quality metrics.

The content of the Morphbank database can be searched by a simple keyword search, by a structured search, or by browsing. The structured search interface includes the following fields: taxon, specimen, view, and locality. In addition, the repository can be browsed alphabetically or hierarchically by a taxon name, an image or specimen identification, a view angle, a locality, or a collection. The logs suggested that Morphbank users might use browsing by a taxon and structured searches more often than simple keyword searches (42 percent and 39 percent versus 20 percent).

In simple keyword searches, users searched the Morphbank repository by (1) scientific name; (2) common name (bee, fish, sunflower, bumblebee, mushroom, etc.); (3) collector name; (4) specimen identification; (5) location; (6) holding institution (Chicago Botanical Garden); (7) taxon part name; (8) habitat type (e.g., water plants); (9) characteristic, or a combination of characteristics facets (e.g., life span, queen bee, flat fish); (10) condition of a specimen part (e.g., swollen members); and, (11) similarity (e.g., looks like a hummingbird). In addition, some users searched for general factual knowledge about a particular species (e.g., bee facts). Interestingly, in spite of this diversity in search types, a search by scientific name was still the dominant search type. Morphbank users searched by a scientific name almost 10 times more often than by a common name.

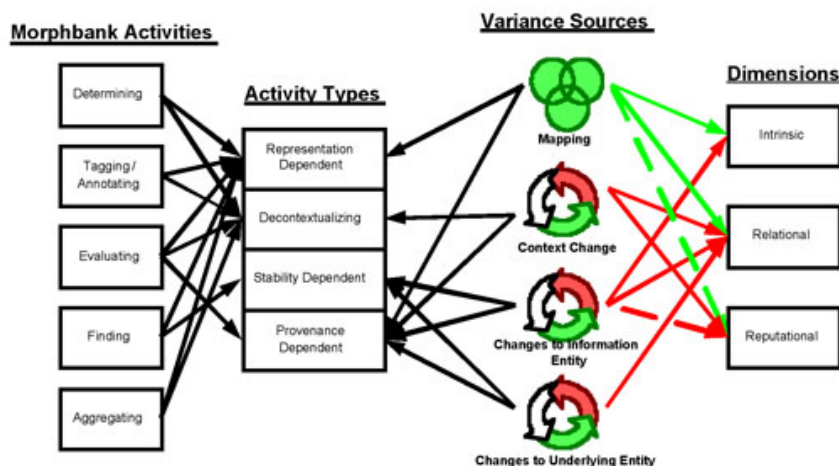
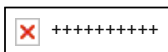


Figure 3: Ontologies, quality measurement, activity theory.



Constructing a quality evaluation model

Thus, the analysis of the morphbank activity system identified the following activities: determining, marking or tagging, evaluating, finding, and aggregating. All five activities may use the ontology as a tool, and some of them (determining, tagging, aggregating) may even modify it by adding, deleting, or editing concepts or classes, definitions, and relationships. The activities represent all four types of activities from the framework and can be prone to all three kinds of problems: intrinsic, relational, and reputational. Furthermore, the activities can be affected by dynamic quality problems (see [Figure 3](#)).

The Morphbank annotations showed the importance of having complete descriptions of both a specimen and a taxon in making a determination. The analysis also suggested that the characteristics of a taxon may vary throughout its life cycle and location, and the ontology may need to represent that.

At present, editing processes in Morphbank are restricted to qualified researchers and the Morphbank project staff. However, the quality of the ontology can also be influenced by changes made from outside the Morphbank community and by changes in the overall context. Taxa may split or merge over time. The ontology continuously needs to be aligned with the changing state of knowledge. Some of the temporal quality variance can be predicted by constructing models of the change in the underlying entity (*e.g.*, Stvilia, 2007). Clearly, predicting this kind of change would be easier for ontologies associated with research involving a systematic or guided hybridization of species than for other kinds of research.

The Web server log analysis suggested that some end users may use common terms when searching for taxon information. Users may also search by keywords, which could be a combination of the instances of two or more classes or facets (*e.g.*, water plants). Hence, the ontology needs to be robust to changes in cultural or community contexts, including vocabularies. Robustness means insensitivity to the variances in factors that cannot be controlled. The largest part of the variance comes from human factors, such as different user needs, which clearly cannot be controlled by ontology creators or designers. A straightforward approach to this problem would be to make the ontology comprehensive to cover as much of a spectrum of the user needs as possible, although this can be against local microeconomic incentives. One could also increase redundancy or parallelism in the content of the ontology to reduce the chances of an activity failure caused by spelling errors or differences in vocabularies used.

Finally, the log analysis found that the authority or reputation of a reference source or key or that of a researcher could serve as a helpful heuristic in evaluating the quality (certainty, in this case) of description or determination of a taxon.

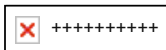
Thus, a quality evaluation model for the morphbank ontology needs to measure variance in the completeness, accuracy, and consistency of the ontology in relation to general cultural and communal reference sources, as well as in relation to the context of a specific activity. The model needs metrics for evaluating the stability and volatility of the ontology — its sensitivity to context and underlying entity changes. In addition, the model needs to evaluate the availability of the ontology's mediation records and the ability to restore a specific state (*i.e.*, provide versioning). This feature can be important in supporting provenance-dependent activities (*e.g.*, morphological phylogenetic activities).

[Table 1](#) lists the IQ dimensions and related metrics for inclusion in the model. Which dimensions to operationalize and which metrics to use will depend on the specific needs and priorities for quality and the costs of metrics.

| Table 1: Quality dimensions and metrics for the quality evaluation model. | | | |
|--|--|---|-------------|
| Dimension | Definition | Metrics | Cost |
| 1. <i>Accuracy/validity</i> | The extent to which information is legitimate or valid according to some stable reference source, such | Spelling error rate = (number of misspelled terms)/(number of terms). | Automatic |

| | | | |
|----------------------------------|--|--|---------------|
| | as a dictionary, or set of domain constraints and norms (soundness), or both. | | |
| 2. <i>Cohesiveness</i> | The extent to which the content of an ontology is focused on one topic. | $\text{AverageICF} = \frac{\sum_{i=1}^n (-\log_2 cf(i) / N)}{n}$ <p>where AverageICF stands for Inverse Class Frequency modeled after the Inverse Document Frequency (Salton and McGill, 1982), n is the number of terms in the class, cf(i) is the number of classes containing the ith item, and N is the total number of classes in the ontology.</p> | Automatic |
| 3. <i>Complexity</i> | The extent of cognitive complexity of an ontology measured by some index or indices. | Cyclomatic complexity of the ontology class graph: number of classes, average fan-out per class. | Automatic |
| 4. <i>Semantic consistency</i> | The extent of consistency in using the same values (vocabulary control) or elements to convey the same concepts and meanings in an ontology. | (Number of inconsistently used elements) / (number of elements). | Semiautomatic |
| 5. <i>Structural consistency</i> | The extent to which similar elements of an ontology (classes, properties) are represented with the same structure, format, and precision. | (Number of inconsistently structured elements) / (number of elements). | Semiautomatic |
| 6. <i>Currency</i> | The currency of an ontology. | Average class currency. | Automatic |
| 7. <i>Redundancy</i> | The amount of non-informative content. | Average info-noise (the size of the informative content, measured in unique word terms, to the overall size of a class definition). | Automatic |
| 8. <i>Naturalness</i> | The extent to which class and | (Number of terms not found in WordNet) / (total number of terms). | Automatic |


| | | | |
|----------------------------------|--|---|---------------|
| | property names and definitions are expressed by conventional, typified terms and forms according to some general-purpose reference source. | | |
| 9. <i>Precision/completeness</i> | Average granularity or precision of the ontology's class model. | Average depth of class. | Automatic |
| 10. <i>Verifiability</i> | The extent to which the correctness of content of an ontology is verifiable or provable in the context of a particular activity. | A ratio of terms and/or assertions supported by or linked to a reference source (ontologies, encyclopedia, research data, and publications) in the total number of terms. | Semiautomatic |
| 11. <i>Volatility</i> | The amount of time the content of an ontology remains valid. | Average update rate. | Automatic |
| 12. <i>Authority</i> | The degree of reputation of an ontology in a given community or culture. | PageRank. | Automatic |



Conclusion

With an increasing number of ontologies available from different sources, and with different models of curation, the ability to evaluate the quality of these ontologies in a systematic way becomes essential for quality-based selection, reuse, and maintenance. Using the activity theory framework of IQ assessment developed in a previous study, we conducted a content analysis of image annotations and search logs of the morphbank database and proposed a general model of quality evaluation for an ontology.

In future research, we will develop a repository of type-specific quality evaluation models for biodiversity ontologies. The models will be ontology type-specific (*e.g.*, plant, fly, etc.)

configurable templates linked to reusable libraries of IQ metrics implemented as computer codes. 

About the author

Besiki Stvilia is Assistant Professor in the College of Information at Florida State University.

Web: <http://mailer.fsu.edu/~bstvilia/>

E-mail: bstvilia [at] fsu [dot] edu

References

- K. Bailey, 1986. "Philosophical foundations of sociological measurement: A note on the three level model," *Quality and Quantity*, volume 20, pp. 327–337.
- A. Burton-Jones, V. Storey, V. Sugumaran, and P. Ahluwalia, 2005. "A semiotic metrics suite for assessing the quality of ontologies," *Data & Knowledge Engineering*, volume 55, pp. 84–102.
- A. Chapman, 2005. *Principles of data quality*. Copenhagen, Denmark: Global Biodiversity Information Facility.
- O. Corcho, M. Fernández-López, and A. Gómez#150;Pérez, 2003. "Methodologies, tools and languages for building ontologies: Where is their meeting point?" *Data & Knowledge Engineering*, volume 46, pp. 41–64.
- S. Crawford, J. Hurd, and A. Weller, 1996. *From print to electronic: The transformation of scientific communication*. Medford, N.J.: Information Today.
- Y. Engeström, 1990. "When is a tool? Multiple meanings of artifacts in human activity," In: Y. Engeström (editor). *Learning, working and imagining: Twelve studies in activity theory*. Helsinki, Finland: Orient-Konsutit Oy, pp. 171–195.
- U. Gasser, 2003. "Information quality and the law, or, how to catch a difficult horse," *Berkman Center Research Publication*, number 2003–08, at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=487945, accessed 6 December 2007.
- T. Gruber, 1995. "Toward principles for the design of ontologies used for knowledge sharing," *International Journal of Human-Computer Studies*, volume 43, pp. 907–928.
- T. Gruber, 1993. "A translation approach to portable ontology specification," *Knowledge Acquisition*, volume 5, pp. 99–220.
- T. Hodkinson and J. Parnell, 2007. "Introduction to the systematics of species rich groups," In: T. Hodkinson and J. Parnell (editors). *Reconstructing the Tree of Life: Taxonomy and systematics of species rich taxa*. Boca Raton, Fla.: CRC Press, pp. 3–20.
- M. Jones, C. Berkley, J. Bojilova, and M. Schildhauer, 2001. "Managing scientific metadata," *IEEE Internet Computing*, volume 5, pp. 59–68.
- J. Juran, 1992. *Juran on quality by design: The new steps for planning quality into goods and services*. New York: Free Press.
- J. Köhler, K. Munn, A. Rüegg, A. Skusa and B. Smith, 2006. "Quality control for terms and definitions in ontologies and taxonomies," *BMC Bioinformatics*, volume 7, number 212, at <http://www.biomedcentral.com/1471-2105/7/212>, accessed 6 December 2007.
- C. Lagoze, D. Krafft, T. Cornwell, N. Dushay, D. Eckstrom and J. Saylor, 2006. "Metadata aggregation and 'automated digital libraries': A retrospective on the NSDL experience," In: *Opening information horizons: Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries, June 11–15, 2006, Chapel Hill, NC, USA: JCDL 2006*. New York: ACM Press, pp. 230–239.
- A. Leont'ev, 1978. *Activity, consciousness, personality*. Englewood Cliffs, N.J.: Prentice-Hall.

G. Leroy and H. Chen, 2005. "GeneScene: An ontology-enhanced integration of linguistic and co-occurrence based relations in biomedical text," *Journal of the American Society for Information Science and Technology*, volume 56, pp. 457-468.

H. Lewen, K. Supekar, N. Noy, and M. Musen, 2006. "Topic-specific trust and open rating systems: An approach for ontology evaluation," *Workshop on Evaluation of Ontologies for the Web (EON2006) at the 15th International World Wide Web Conference (WWW 2006)*, at <http://smi.stanford.edu/smi-web/reports/SMI-2006-1142.pdf>, accessed 6 December 2007.

B. Ludäscher, I. Altintas, C. Berkley, D. Higgins, E. Jaeger, M. Jones, E. Lee, J. Tao and Y. Zhao, 2006. "Scientific workflow management and the Kepler system," *Concurrency and Computation: Practice and Experience*, volume 18, pp. 1039-1065.

A. Orme, H. Yao and L. Eitzkorn, 2007. "Indicating ontology data quality, stability, and completeness throughout ontology evolution," *Journal of Software Maintenance and Evolution: Research and Practice*, volume 19, pp. 49-75.

C. Parr, A. Parafiynyk, J. Sachs, L. Ding, S. Dombush, T. Finin, D. Wang, and A. Hollander, 2006. "Integrating ecoinformatics resources on the Semantic Web," *Proceedings of the 15th International Conference on World Wide Web*, pp. 1073-1074.

G. Salton and M. McGill, 1982. *Introduction to modern information retrieval*. New York: McGraw-Hill.

D. Strong, Y. Lee, and R. Wang, 1997. "Data quality in context," *Communications of the ACM*, volume 40, pp. 103-110.

B. Stvilia, 2007. "A model for information quality change," In: M. Robbert, R. O'Hare, M. Markus, and B. Klein (editors). *Proceedings of the 12th International Conference on Information Quality (ICIQ2007)*. Cambridge, Mass.: MIT, pp. 39-49.

B. Stvilia, 2006. "Measuring information quality," Ph. D. dissertation, University of Illinois at Urbana-Champaign, at <http://wwwlib.umi.com/dissertations/fullcit/3223727>, accessed 6 December 2007.

H. Van de Sompel, C. Lagoze, J. Bekaert, X. Liu, S. Payette, S. Warner, 2006. "An interoperable fabric for scholarly value chains," *D-Lib Magazine*, volume 12, number 10, at <http://www.dlib.org/dlib/october06/vandesompel/10vandesompel.html>, accessed 6 December 2007.

Editorial history

Paper received 31 August 2007; accepted 15 November 2007.



A model for ontology quality evaluation by Besiki Stvilia is licensed under a [Creative Commons Attribution-Noncommercial 3.0 United States License](http://creativecommons.org/licenses/by-nc/3.0/)

A model for ontology quality evaluation by Besiki Stvilia
First Monday, Volume 12 Number 12 - 3 December 2007

<http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/rt/prINTERfriendly/2043/1905>