A Classification of Models

Abstract: Libraries, bibliographic databases, and the WWW/Internet do not provide access to a coherent, unified class of works known as models. This paper presents a framework for the cataloging and classification of one class of models, namely scientific models. Scientific models are defined as works and models classification is based on facet analysis.

1. Introduction

Scientific models can be used to improve science learning and integrate scientific research with teaching. NASA (1988) summarizes the kinds of prerequisites that models useful for science learning, specifically geo-sciences, must enable. These include:

1. Acquisition of observations - the model must exhibit a selective attitude to information;
2. Analysis and interpretation of the observational data - the model must be structured, pattern-seeking and replicating;
3. Construction of and experimentation with conceptual and numerical models; and
4. Verification of the models - reliability and validity for further testing, experimentation, and replication.

ThinkerTools (2001) at the University of California at Berkeley has developed Morton Modeler, a computer agent, who walks students through the process of designing and building good scientific models. Projects such as the Alexandria Digital Library are trying to include models for providing learning spaces in digital libraries (Coleman, 2001). A metadata standard with approximately 150 elements of description has been developed for describing computational models in the geo-sciences (Hill, 2001).

2. Background

The word model has many definitions. In ordinary, everyday usage it can be used to indicate a three-dimensional reproduction, usually on a smaller scale, a design or style of structure, a person or thing that is worthy of imitation, a person hired to pose for an artist or display clothes, etc. (OED Online, 2001). Model can also be used as an adjective in the sense of exemplary.

In the sciences and social sciences, the word model has different connotations and this study limits its analysis of models to the sciences. Minshull (1975) argues that a model in geography can be a theory, law, hypothesis, structured idea, a role, relation, equation, reasoning, or synthesis of data. Representational objects such as maps and satellite images are also considered models (Chorley and Haggett, 1967). The increasing use of information technology and the advent of the World Wide Web (WWW) has added another important
dimension to the scientific modeling enterprise. Many models are now available as black box models and web applets. Black box models conceal complex computational and other features of modeling while applets typically do not require special computers to run the models. Visualizations, simulations and datasets are also increasingly being referred to as models; sometimes these are products (outputs) and inputs of the activity of scientific modeling.

Models are the intellectual artifacts, the creative products of scientific research and important tools for learning science skills. (Heckhausen, 1972, 85). Yet libraries, bibliographic databases, and the Internet/WWW do not provide collocated and easy access to models. A partial list of problems in the information retrieval of models in one area, constitutive relations in Civil Engineering, includes:

1. Vocabularies used currently in the major databases for the discipline do not have or use a descriptor constitutive models.

2. There is a lack of terminological consistency across different sources. Information resources and packages such as books, theses, dissertations, and reports, about models are described in library catalogs using the controlled vocabulary scheme of Library of Congress Subject Headings (LCSH). Smaller information packages such as articles from journals and conference proceedings are covered by indexing and abstracting sources and bibliographic databases. For these, the preferred controlled vocabulary varies from database to database. For example, COMPENDEX is one of the major databases in engineering and it uses its own scheme. Therefore, users must become familiar with at least two different vocabularies to search for constitutive models in the library catalog and a bibliographic database or periodical index.

3. Unmatched knowledge structures: Vocabularies map the knowledge of experts in a discipline; this is usually the authors and writers of the articles indexed and based on literary warrant. The knowledge structure thus derived is often significantly different from practitioners and learners who may have incomplete knowledge structures. An informal survey of Civil Engineering faculty at the University of Arizona at Tucson showed that for improving learning constitutive models as one cohesive class under which items with different facets/aspects about constitutive modeling could be further sub-arranged or displayed was preferred. For example, rather than specific subjects under constitutive models such as creep, stress and strain, faculty would have preferred model facets/aspects such as a typology (elastic, plasticity-based, discontinuity, novel) and object modeled (tunnel, slope, etc.).


Thus, the overarching question can be stated quite simply, what is a scientific model and how can it be represented in the library catalog to reveal disciplinary structures and facilitate information retrieval for learning?

3. Research Questions

This models classification project is part of a larger study that is developing a
registry and prototype database/catalog of scientific models. Findings for the research questions that are explored in this paper are:

1. Are scientific models works?
2. What are the facets in a scientific models classification scheme?

Other subsequent research questions that the study will be exploring include:

1. What is a models classification scheme?
2. What are characteristic descriptive properties of models that can be automatically derived as a function of model classification?
3. How can such a classification/nomenclature be used in 1) retrieving information about models, 2) archiving, and 3) model execution via the WWW browser.

4. Hypothesis

It is hypothesized that a scientific model is a work. It is further hypothesized that classification or nomenclature for models can serve best as the basis of models description and retrieval in digital libraries with learning spaces. Classification is making a re-emergence as a viable method for the semi-automatic organization of electronic resources. See for example, Project DESIRE (Koch, 1997) and Project Scorpion (The Scorpion project, 2002). Because scientific models are complex objects that have multiple representations, exist in heterogeneous forms and formats, and include interdisciplinary problems and domains, classification may be useful in improving retrieval of models by revealing implicit knowledge structures, subject and other relationships.

5. Methods

A preliminary study, to test the notion of scientific models as works, was conducted. It included document, content analysis, as well as retrieval of selected scientific models in a bibliographic utility, OPAC, controlled vocabulary and classification schemes, and bibliographic databases. Methods and preliminary results are fully documented elsewhere and readers are encouraged to refer to it or contact author for details (Coleman, in press).

6. Results

In the interest of brevity only results from two specific parts of the study are presented here. First are the findings from the document analysis regarding the physical composition of one scientific model, Atmosphere-Ocean Model from NASA at the URL: http://aom.giss.nasa.gov/. Scientific models appear to have the following forms and formats:

1. Theory or Hypothesis (text file or graphic diagram/image)
2. Observations and Measurements (data sets in various formats)
3. Computer Hardware (descriptions only)
4. Computer Software (this includes actual software executable code, other software essential for running the model code itself, documentation (technical and user guide that accompanies the software)
5. Animations (images, video)
6. Visualizations (images, video)
8. Reports, Reviews, Annotations, Experiments, Articles (text)
9. Tools (applets, services)
10. Mathematic (algorithmic, numeric notation)
11. FAQs, Mailing Lists/Listservs, Discussion Groups, Events (various).

Second, a summary of the findings of the content analysis of published literature (texts) shows that the following facets are present in models in the area of water quality:
1. Concept: is an idea, the traditional subject (for example, calculus of variations)
2. Object: the object studied in the model
3. Discipline: the major discipline to which this model belongs (may be determined either through author affiliations or other means)
4. Phenomenon: the phenomenon being modeled
5. Process: the process being modeled
6. MathRepresentation: the mathematical functions, equations used
7. Software: the software needed to run the model
8. FundamentalLaw: the fundamental laws that the model is based upon
9. Type: the type of model based on its purpose
10. Variable: number, types, conditions, and variables in this model
11. Problem: the problem the model is analyzing stated often as a question
12. Theory: is there an existing theory or research group to which this model belongs

7. Discussion

Drawing from information management traditions in libraries, archives, and museums, there are at least three perspectives that can be used for organizing scientific models as works. The library tradition emphasizes collection of individual objects, the independent bibliographic unit, or the individually available information package. If this framework were to be used, the emphasis would be on descriptive cataloging. The archival tradition emphasizes collections of related objects. The museum tradition emphasizes consistent, constant, change. Disparate objects are often collected as exhibits based on changing or newly revealed and articulated relationships.

Examining scientific models as works provides a strategy that appears to draw the best from all three traditions noted above. It is described as a series of practical steps in the development of a classification scheme for models.

Step 1: “A work is the intellectual content of a bibliographic entity; any work has two properties: a) the propositions expressed, which form ideational content and b) the expressions of those propositions (usually a particular set of linguistic (musical, etc.) strings) which form semantic content.” (Smiraglia, 2001). Semantic content in scientific models includes mathematical expressions, formal propositions and hypotheses, and statements of laws. Ideational content includes ideas about objects, processes, and relationships, usually within or for specified spatial and temporal scales, and formally, semantically expressed as mathematical equations and algorithmic notation. The ideas include both observables (verified and expressed as measurements) and non-observables (hypothetical data, mathematical equations).

Step 2: A work is a bibliographic entity. Scientific models as bibliographic
entities have two properties: physical and conceptual.

Step 3: The physical components of a scientific model can be determined in terms of its form (what the instantiation is) and the following forms and formats of models exist

1. Textual works – includes articles, abstracts, bibliographies, reviews, analysis, software documentation.
2. Datasets – includes observations and measurements of the observed phenomenon, object, process reported as data, images, visualizations, and graphs.
3. Software – includes computer code, both source code and downloadable executables.
4. Services – includes interactive and other services (animation applets, databases, indexes, contact pages, submit forms, etc.)

When conceptual components are examined the facets, the basic ideas, the model expresses, can be abstracted. Even more than just the ideas, the ideational (subject + other) relationships are important in modeling. Conceptual components are called model concepts and relationships and the 12 facets from the results section (see section 6 above) are the preliminary facets for subsequent development.

Step 4: We are developing a prototype catalog of scientific models using this preliminary classification. We are using the Dublin Core Metadata Element Set, Version 1.1, as the base metadata scheme for cataloging selected models. In addition, the 12 models facets are being framed for the area of water quality with other general additional facets such as form, time, space, author/group and bibliographic relations. A prototype database should be available by the end of the year 2002. A sample framework is shown below.

<table>
<thead>
<tr>
<th>A Models</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Scientific models</td>
<td>rasps</td>
</tr>
<tr>
<td>(by object)</td>
<td>5 (by mathematical representation)</td>
</tr>
<tr>
<td>2 water</td>
<td>51 differential equations</td>
</tr>
<tr>
<td>21 rivers</td>
<td>6 (by software)</td>
</tr>
<tr>
<td>22 lakes</td>
<td>use MIME formats here</td>
</tr>
<tr>
<td>3 (by process)</td>
<td>7 (by fundamental law)</td>
</tr>
<tr>
<td>infiltration</td>
<td>8 (by discipline)</td>
</tr>
<tr>
<td>4 (by phenomenon)</td>
<td>9 (by concept) – use traditional subject headings here</td>
</tr>
</tbody>
</table>

Step 5: Each unit of analysis is given a class number for each facet in addition to the metadata that is created for the unit. Most of the facets are currently using controlled values from different schemes until the classification framework is completed.

8. Conclusion

Scientific modeling conceals an incredible amount of intellectual relationships that traditional bibliographical tools, primarily the catalog and the index, neither capture nor describe from the texts, documents, and items about models. Our increasing awareness of conceptual and textual instability of electronic forms requires active investigation and experimentation with other types of knowledge.
organization and representation structures. This paper has presented an alternative solution to view scientific models as *works* and use a models faceted classification scheme for their subsequent display and information retrieval. Decades of research both in information retrieval and information seeking behavior complemented by the widespread success of Internet search engines has shown us that users tend to disregard Boolean searches, human indexing as opposed to machine indexing does not improve search performance significantly, and that users want a few relevant, good materials. Assessing relevance in terms of disciplinary structures has never been researched. Therefore the subsequent use study of the developed models catalog/database and the evaluation of retrieval and displays based on classification should reveal interesting findings.

**Acknowledgements**

The comments of three anonymous reviewers who approved the funding for the models classification study is gratefully acknowledged. Feedback from the anonymous reviewers of the Seventh International ISKO Conference review panel were also very useful and I thank them. I thank my former student, Olha Buchel, for many meaningful discussions on the cataloging and classification of models and for her support in classification, database, and catalog design. Thanks to University of Arizona Professor Muniram Budhu, Dept. of Civil Engineering, Dr. William Rasmussen, Dept. of Agriculture and Hydrology, and Dr. John Kemeny, Dept. of Rock and Mining Engineering for their participation in the informal survey of models for teaching and learning engineering concepts. This study is supported by the University of Arizona, Office of the Vice President for Research Faculty Small Grant, 2001-2002

**References**


