Anita S. Coleman
University of Arizona at Tucson

A Classification of Models

Abstract:
Scientific models are critical, analytical tools and objects that can be included in digital libraries. This paper presents a preliminary classification scheme for the cataloging of scientific models. Scientific models are defined as *works* and the classification is based on facet analysis and corroborated by an user study.

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 in learning 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 (structured, pattern-seeking and replicating);
3. Construction of and experimentation with conceptual and numerical models; and
4. Verification of the models, together with their use to furnish statistical predictions of future trends (testing, experimentation and replication).

Projects such as the Alexandria Digital Library suggest that models are worthwhile objects for providing learning spaces in digital libraries (Coleman, 2001). A metadata standard with approximately 150 elements of description has been developed for describing computational, geographical models (Hill, 2001).

2. Problem:

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. 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 can be a theory, law, hypothesis, structured idea, a role, relation, equation, reasoning, or synthesis of data. Maps and satellite images are also models. How can digital libraries define, collect, organize, and describe these complex digital objects?

3. Hypothesis:

It is hypothesized that a classification or nomenclature for models can serve best as the basis of models description and retrieval in digital libraries. Classification is making a re-emergence as a viable method for the semi-automatic organization of electronic resources (Project DESIRE, Project Scorpion). Further, models are analytical

tools that can be shared by the sciences and a fruitful area for the investigation of universals. Geography is traditionally a discipline of synthesis and therefore provides an excellent corpus for identifying a models classification that would be extensible to the physical sciences. Geography also has provided a classification of models, notably through the work of Chorley (1967). Finally, time and space are well-developed categories in geography.

4. Research Questions:

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.

5. Project Description:

A definition of models in geography is currently under development based on document analysis (textbooks, curriculum materials and thesauri). It is being conducted in two phases. In the first phase about 500 objects collected from textbooks and curriculum materials, where they are referred to as models, are being analyzed. A preliminary definition has identified at least two fundamental purposes of models, predictive and descriptive. Three more categories can be identified as components of one type of model, a simulation model system: conceptual, mathematical, and physical. The components produce data depending on the phase or stage in the model development process. These are briefly explained below:

1. Research foci are the topics expressed in terms of geographic processes and form the basis of the conceptual sub model.
2. A mathematical model is another sub model (part of a model system)
3. The simplest mathematical sub model needs at least three types of variables and a set of operating system characteristics linking them. The three sets of variables are: input variables, status variables (the internal mathematical constant), and the output variables (which depends on both input and status variables).
4. If the model is predictive cause and effect must be clearly distinguished. This is not so important in a conceptual descriptive model (only explains, doesn't predict nor is used in planning).
5. Model strategy is the fitting and testing of the model by choosing the type of mathematical operations most suitable to the type of system one is trying to model.
6. There is testing the model to see if it works as planned before using it for prediction.
7. Reporting or Publishing the model.

A document analysis of the physical composition of simulation model systems (Atmosphere-Ocean Model from NASA at the URL: http://aom.giss.nasa.gov/) has

shown that models are complex digital objects (also known as distributed collections or distributed physical packages). They 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)
7. Reports, Reviews, Annotations, Experiments, Articles (text)
8. Tools (applets)
9. Mathematic (algorithmic, numeric notation)
10. FAQs, Mailing Lists/Listservs, Discussion Groups, Events (various).

The second part of the study is analyzing the representation of models in five library tools (Water Resources Abstracts, GeoRef, Gemet, Library of Congress Subject Headings, Dewey Decimal Classification). Using the fundamental categories of Ranganthan's colon classification, Time, Space, Matter, Energy, Personality and borrowing from existing classification schemes in Math (American Mathematical Society), Digital Formats (MIME), and learning (ERIC), a classification scheme for one small area, water models is being developed. A field test with users is planned using a small prototype database and modifications will be incorporated into the scheme. This classification scheme for models it is hoped will be a feasible method whereby models can be more easily added to library collections.

Notes and References:

Chorley, Richard J. and Peter Haggett. 1967. Models in Geography. London: Oxford University Press.

Coleman, A.S., et al. 2001. Learning Spaces in Digital Libraries. Lecture Notes in Computer Science, Vol. 2163. Proceedings of ECDL 2001, the 5th European Conference on Advanced Research and Technology for Digital Libraries, Darmstadt, Germany, September 4-9, 2001.

Hill, L. et al. 2001. A Content Standard for Computational Models. D-Lib Magazine. URL: http://www.dlib.org/dlib/june01/hill/06hill.html

Minshull, Roger. 1975. An Introduction To Models In Geography. London, Longman.

NASA. 1988. Earth System Science: A Closer View. Washington D.C., NASA.