

Four interacting aspects of a geological survey knowledge system

T.V. Loudon

British Geological Survey, West Mains Road, Edinburgh EH9 3LA, United Kingdom.

email: tv1@bgs.ac.uk

Postprint of article published in *Computers & Geosciences*, 2009.

Bibliographical reference: Loudon, T.V., 2009. Four interacting aspects of a geological survey knowledge system. *Computers & Geosciences* **35**, 700-705
doi:10.1016/j.cageo.2007.12.009

Abstract

The developing cyberinfrastructure affects the knowledge system by which geological surveys collect, represent, and communicate their knowledge, and thereby influences their view of the geology. Consequences for four interacting aspects of the overall system (infrastructure, business models, geological framework, and surveying methods) are outlined. Although each reflects a different area of expertise, all aspects must work together to support an incipient change of emphasis in survey work – from publishing maps and supporting documents, to contributing to a whole-Earth knowledge system that responds flexibly to user needs.

Keywords: cyberinfrastructure; surveying methods; business models; Earth systems; interoperability.

1. Introduction

The future directions of the geoscience knowledge system and therefore of geological survey organizations will be influenced by many external forces. Of these, the developing network of support services based on computing, information and communications technology (the cyberinfrastructure) is particularly noteworthy. It affects the way surveyors collect, represent and communicate their ideas, and in consequence their view of the geology. The surveying process takes place where knowledge of the solid Earth is tested, enhanced and extended by relating it to its real-world counterpart. The core of that knowledge is at present represented largely by lithostratigraphic maps and associated documents. Potentially, it will be supported by a cyberinfrastructure that will provide the mechanisms for knowledge representation and communication, along with the computing power to model solid-Earth systems as a component of a global knowledge system.

Consequences for four aspects of the system, each calling for different types of expertise, are outlined here. They can be described by models (in the sense of conceptual constructs representing a simplified view of reality for a particular purpose) and their interactions. They are: the infrastructure model (how knowledge is represented, stored and communicated); the business model (why surveys are undertaken); the geological framework model (how geological information is organized); and the surveying methods model (how the survey is carried out). Consideration of all these aspects is required for practical forward planning.

2. Cyberinfrastructure model

The model of the cyberinfrastructure represents the structure, mechanisms and facilities to store, process and communicate representations of global aspects of knowledge. The physical manifestations of geological records are regarded in this paper as an infrastructure component (Fig. 1). Cyberinfrastructure design is primarily the concern of 'e-scientists'. But the geological community has a responsibility to respond to its development, and to ensure that the design takes its specific requirements into account.

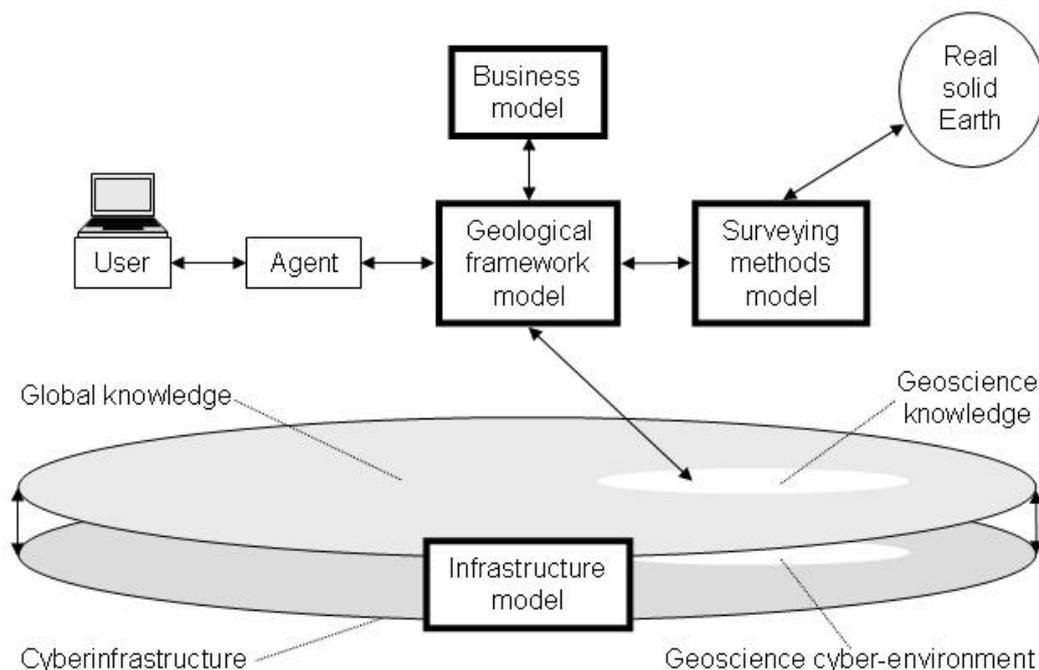


Figure 1: Relationships of four models in a survey knowledge system, as explained in the text. Arrows show information flow.

The cyberinfrastructure is reacting to exponential growth in computing power, communications bandwidth and storage capacity, as the Internet acquires features of the semantic Grid¹. The Grid (Foster and Kesselman, 2003) is seen as supplying information (representing aspects of knowledge) as a commodity, and the semantic aspect implies an interest in incorporating interpretative reasoning processes. Unrestricted by artificial boundaries of geography or discipline, the Grid should support the worldwide extent of geology, and strengthen the links to other disciplines. Many geological processes and events of particular concern to mankind involve interactions among the lithosphere, hydrosphere, atmosphere and biosphere. Their separate disciplinary approaches can be unified in the concepts of Earth systems science (National Research Council, 1993), and supported by a formal framework and exploratory apparatus² from computer science, are appropriate for Grid implementation.

The semantic Grid should thus enable knowledge from many fields to be readily brought to bear on geological issues, and vice versa. Shared standards, concepts, ontologies and interoperable models facilitate collaborative work across related fields of study. They should make it possible to combine information and software from many sources, and make it available to a wide range of application fields, using methods such as the Service-Oriented Knowledge Utility³. In response, some geological surveys are changing the emphasis of their business model⁴: away from publishing geological maps and related documents; towards

¹ Semantic Grid Community Portal. <http://www.semanticgrid.org/>

² Emmott, S. (Ed.) 2006. Towards 2020 Science.

<http://research.microsoft.com/towards2020science/downloads.htm> and <http://www.nature.com/nature/focus/futurecomputing/index.html>

³ Next Generation GRIDs Expert Group Report 3, 2006. Future for European Grids: GRIDs and Service Oriented Knowledge Utilities. ftp://ftp.cordis.lu/pub/ist/docs/grids/ngg3_eg_final.pdf

⁴ Natural Resources Canada, 2006. Consolidating Canada's Geoscience Knowledge http://ess.nrcan.gc.ca/2002_2006/ccgk/index_e.php

contributing to a whole-Earth knowledge system, from which users can obtain flexible responses to match their specific needs.

3. Business models

Business models that set out the objectives of a geological investigation are primarily the concern of geological management. They arise in many diverse activities, for example, exploration and exploitation of resources (such as water, hydrocarbons and minerals), civil engineering, hazard assessment, research, education, or just satisfying curiosity. Each business model, and each project within it, has its own implicit or explicit profile, identifying which geological aspects and properties are of salient importance in achieving the objectives. Profiles of this kind can guide the surveying process, and can also help to evaluate the relevance of the results for other applications. For example, superficial deposits might be known from business models to be of interest to a civil engineering study, but not to an oil exploration study. The profiles can be recorded as part of the provenance of information, adding value by deepening the users' understanding of the results. The business model does not necessarily involve economic considerations, although its concern with objectives suggests an interest in a shared metric of utility (such as money). It may include procedures for acquiring and allocating resources in order to develop and deliver the desired products, and may involve cost recovery and perhaps profit.

A primary objective of the business model of a geological survey organization is likely to be the development, recording and communication of an authoritative, coherent account of the geology of a region. The objective is generally achieved by reviewing what is already known; systematically observing and measuring salient geoscience properties in the field; analyzing and interpreting the results in the context of the rock types they characterize; and evaluating, integrating, summarizing, and communicating the findings. The business model therefore interacts with the cyberinfrastructure model, as well as the interlinked knowledge described in the framework model, and the methods implied by the surveying model. The emphasis is generally on supplying basic geological information, rooted in an understanding of the nature, distribution, history and configuration of the rock types, to support a wide range of commercial, regulatory and research activities. The business model therefore seeks to provide information of wide relevance, providing a quality-assessed core of geological knowledge on which many internal and external applications can build, and to which they can relate their results. The objectives will remain, but the means of achieving them will surely change.

Current business models will require review in the light of Grid developments. The value of information to users can be maintained only by keeping pace with the evolving cyberinfrastructure and providing appropriate, high-quality information. Its usefulness depends on its accuracy (implying quality assessment), relevance to the actual and potential requirements (implying accurate classification and linkage), and the ease with which it can be accessed, used and understood (implying standards, appropriate technology and interoperability). Even within a survey, there is a hierarchy of business models at various levels of detail, including the objectives hidden in the tacit knowledge of individuals and in implicit project objectives, which must be reconciled with the overall geological survey business model. The shared business model adds greatly to the practical value of the final products. The information is evaluated to assess its suitability for the permanent record, in survey organizations by internal quality assessment and in external publications by editorial judgment assisted by reviewers. Citations in the literature reflect subsequent evaluation by users, which includes further verification by testing against the real world.

Many survey organizations have an important secondary role in storing and disseminating information from external sources, such as civil-engineering borehole descriptions or results of academic studies, collected according to different business models. The increasing flexibility and diversity of a survey's hypermedia information store will lead to closer interaction with a wider range of users. This process is supported in some disciplines by on-

line systems designed for user collaboration, such as OpenWetWare⁵ and Open Notebook Science⁶. It can therefore be appropriate for a survey organization to maintain a collaborative information system that includes comments, and external contributions that retain their own business models. Value can be added to external contributions by extending their relevance through links with the core information and sharing standards where appropriate. Geological survey organizations can make a vital contribution by formalizing a framework model and providing its core content. A comprehensive framework should be able to connect items of information at all levels of detail from many sources, indicate their provenance, and help users to reconcile the information with their own needs.

4. A framework model

The geological framework model attempts to depict and clarify the principal relationships among the findings of geology, and their links with applications and with other aspects of science. It is potentially a specialized component of the geoscience cyber-environment in the infrastructure model (Fig. 1), but, as it is primarily the concern of the geological community, it is considered separately in this paper. Frameworks are intended to foster the collaborative nature of science, encouraging individuals and organizations to work together in building knowledge for their mutual benefit. The existing framework for the practice of geology, embedded in the conventions of geological procedures and documentation, is largely taken for granted. Its successor requires explicit consideration to make it accessible to software, with a view to simplifying the users' access to the geological knowledge stored in the infrastructure.

Searching for appropriate information in the diverse distributed stores representing aspects of global knowledge is a daunting task. Many users, including those unaware of the relevance of geology to their requirement, may initially use a standard search engine to search for combinations of pertinent words. It is helpful if surveys provide Web pages that enable search engines to lead users to the appropriate specialized framework required for more detailed search, such as the geological framework model.

The aim of the framework model is to provide a map of the contents of geological knowledge, in order to identify regions that match, or are close to, the user's requirements, which can then be linked to relevant information. Unlike a geographical map, the map of the geological framework has many dimensions. As a starting point for discussion, [Loudon and Laxton \(2007\)](#) suggested basing the contents map on a solid Earth systems model (a model of its systems, in the sense of sets of interacting parts operating as a whole and organized to perform a specific function). The model refers to the three-dimensional disposition and configuration of the present-day objects of geoscience (the location and arrangement of the things of geological interest), to their observed and interpreted properties and composition, and also to their history throughout geological time, including the processes that created and altered them and are critical to the interpretation. In other words, it refers at all levels of detail to what is there, what it is called, where it is, how it is arranged, what its properties are, what it is made of, where it came from, how it got there, and how we know.

The solid Earth systems model links to a second part of the framework, namely relevant ontologies, described by Raskin (2006) as "formal representations of concepts and their mutual relationships in a form that captures domain knowledge". Standard generic ontologies could help to relate geological information to information in related fields. A third essential part of the framework is a set of indexes, which relate regions in the model, along with terms in the ontologies, to information (including conventional representations) held in distributed information stores. Each set of information would ideally include metadata describing its

⁵ OpenWetWare <http://openwetware.org>

⁶ Open Notebook Science <http://drexel-coas-elearning.blogspot.com/2006/09/open-notebook-science.html>

provenance (including its business model) and constraints on its analysis, in a machine-readable form where appropriate. The framework of solid Earth systems model, ontologies and indexes could, on the one hand, help surveyors to organize their information, record the course of their investigations, structure scientific workflows to generate products for users, and establish links to broader aspects of Earth systems science. On the other hand, the framework could help users to locate and display relevant information, evaluate its content, reconcile the fragments and facets of knowledge, and track threads of reasoning. It could also identify where ontologies (specifying how aspects of the model are described) can be shared with other disciplines, thus promoting the interoperability that enables diverse analytical models to work together. It is therefore central to the geological survey knowledge system (Fig. 1).

5. Surveying methods model

The surveying methods model describes the procedures by which information is gathered for adding to or amending the solid Earth systems model, and is a specific responsibility of the surveyors. Surveyors, in the same way as any other user, must start by acquiring relevant information. This can then be tested against observations of the real solid Earth, and amended and extended by new observations or interpretations, which in due course can take their place in the geoscience knowledge recorded in the infrastructure. Only fragments of the solid Earth systems model can be constructed from observation and analysis. The task of surveyors is to select and survey such fragments guided by background knowledge, interpret their significance, and assemble the interpretations as a coherent whole. The cyberinfrastructure enables the surveyors to record explicit connected pathways between conclusions, observations, and the process of reasoning and interpretation at all levels of detail. Compared with conventional methods, these can provide a more transparent and rigorous statement of scientific evidence, can reduce ambiguities across disciplines and application areas, and make results accessible to a wider range of users. The potential consequences include major change in methods of surveying.

The methods are changing as electronic support in the field and office develops, and the costs of storing, processing and communicating information plummet (Brodaric and Gahegan, 2006). The more powerful infrastructure makes it feasible to introduce new concepts into geological surveying. Also, search engines make scientists more aware of methods used in other fields and the prospect of more widely coordinated methods. The methods, which support and enhance each other, could enable survey organizations to build a more rigorous and effective record of regional geology, anchor it more firmly in its multi-disciplinary setting, and provide an improved response to their business model.

Proposed methods include those based on concepts such as: three-dimensional spatial models of geological objects; object-oriented views of instances, classes and inheritance; scale-space to identify the ranges of resolution at which observations were made, objects exist and processes operate; zero-crossings for reproducible selection of object boundaries; Gaussian filters for more natural continuous scale change; emergent processes for more convincing interpretations; hypermedia integration of spatial, narrative and tabular information at all scales or granularities; representation and indexing of narrative threads of evidence, description, reasoning and interpretation; links to remotely sensed images that provide or clarify evidence (such as landforms, outcrops); workflow-based recording of observations, and the methods and reasoning that link them to the interpretation; process-based interpolation methods and deformable inhomogeneous spatial models for space-filling by justifiable and reproducible procedures; records of shape statistics of objects, suited to regional comparisons as they are unaffected by moving about in three dimensions and (within limits) in scale-space; visualization that separately superimposes knowledge of existence and shape on likely position; and Bayesian statistics to track opinions and evaluations of accuracy.

The infrastructure must accommodate legacy information and identify its limitations. But it also offers an opportunity for local experts to rework earlier surveys to take newer approaches into account in order to refurbish the legacy information, add to its value in future information systems, and develop techniques that could lead to more comprehensive digital field support.

6. The knowledge system

The diagram (Fig. 1) shows how the four models just discussed might work together within a geological survey knowledge system. The flow of information is shown by arrows. Users submit requirements to the system and receive results by means of an agent (a piece of software that acts with a degree of autonomy to accomplish tasks on behalf of its user). Through dialog with the user, the agent can clarify the user requirements as a profile of interests, defined in terms of relevant regions of the geological framework model. The user's business model could also help the agent to establish the user's profile of interests, and therefore which sources of information and which properties are of most interest. The agent might also identify relevant material by comparing the user's profile with the provenance metadata of information sources. The profiles identify the relative importance of properties represented in the framework model, which in turn can be related through indexes to appropriate information.

At the simplest level, the user profile could be matched with reusable items (represented as hierarchies of objects), which can represent survey results at any level of detail, from an outcrop observation to a completed map or a regional summary. The framework might also locate reusable pathways, represented by scientific workflows (Ludäscher *et al.*, 2006), describing a sequence of procedures that can be applied over a long period to ever-changing information stores and software sources. For example, the user might wish to select, retrieve, assess, analyze and visualize the most up-to-date geotechnical and associated information for a specified area using the latest display techniques, simply by invoking an existing workflow. The framework can also make it possible to track and compare hypertext threads of reasoning, which might, for example, link a geological interpretation to many sources of supporting information. Based on the profile originating from their business model, surveyors could also acquire relevant existing information, and call up appropriate techniques from the surveying methods model to interface with the solid Earth. Within the framework, the solid Earth systems model could help to organize their investigation and the records of their results.

The organization of geological information is changing. Conventional geological maps represent interpretations of salient aspects of the geology. They enable users to visualize and recognize patterns on the map and relate them to geological explanations. The maps locate observations and interpretations on the ground, where they can be tested against the real world. Digital cartography, geographic information systems and computer visualization provide greater flexibility – compare different aspects of the geology, pan around to compare different areas, zoom in and out to compare features at different scales, and apply geometric transformations to remove extraneous effects, such as removing the results of faulting and folding before studying facies variation. The computer implementations, however, are essentially reductionist methods. They break down the information content into its basic objects, representing observations and items depicted on the map, which can then be reassembled in various ways. But isolated objects have little meaning. Geologists make sense of their observations by integrating them with through reasoning and interpretation, leading to testable conclusions. In response to a similar situation, disciplines such as ecology, biology, and medical science⁷ place increasing emphasis on systems – less emphasis on objects in their own right, more emphasis on the network of connections between objects. The solid Earth systems model also emphasizes the systems view. In the evolving infrastructure, geologists

⁷ European Science Foundation, September 2007. Systems Biology: A Grand Challenge for Europe <http://www.esf.org/publications/medical-sciences.html>

will have an opportunity to represent, record, and communicate the interactions and relationships of information items, and develop their holistic, historical narratives of evolving configurations of geological objects in the context of the environments and geological processes that shaped and changed them. A shared overview of the solid Earth system, seen as part of the framework, could avoid needless repetition and encourage interoperability.

7. Conclusions

Theoretical proposals, such as this one, are an inexpensive means of exploring the design and potential of a future knowledge system. In practical terms, however, geological organizations generally aim to remain in the mainstream of infrastructure developments, and ensure that their resources of knowledge, information and procedures are routinely updated in step with the technology (Howard *et al.*, this volume). Geologists have readily adopted procedures that are analogous to conventional methods, including word processing, digital cartography, illustrated presentations, databases, statistical analysis, bibliographical searches, electronic mail, handling and editing digital manuscripts, geographic information systems and visualization, and the World Wide Web. Mainstream solutions are widely used for discovering relevant material with search engines, delivering digital products for desktop editing and printing, and transferring micro-payments. Such methods, however, are peripheral to the underlying science, and do not address the full potential of the cyberinfrastructure and its impact on geological survey organizations.

Existing technical solutions, as the following examples illustrate, are not fully integrated. Procedures to deliver-and-print at the user's desk cannot handle the size and precision of existing cartographic products, which require specialized printing equipment of limited availability. Several infrastructure models have features relevant to geoscience investigation, such as the visualization and interaction facilities of the 'massively multiplayer online role-playing game', but are incompatible with commonly used geological methods. During a long period of transition, the huge legacy of information dependent on conventional representations and procedures will exist in a different setting from that of the developing geological cyber-environment, and yet they must operate side by side. Only robust and well-supported systems will be able to cope with the diversity and the need to migrate as the solutions evolve.

It has been argued in this paper that various approaches might help geological survey organizations to benefit more fully from a Grid-based approach. These include: organizing information and ideas in modular form within a framework of models for solid Earth systems; developing and evaluating new surveying techniques; and reassessing earlier concepts in the light of such developments. Some established standards for geological classification and nomenclature may be inappropriate to a Grid-based knowledge system – for example, some stratigraphical classifications are tied more strongly to map representation than to scientific explanation. Nevertheless, the developing cyberinfrastructure should in due course lead to a clear, integrated, mainstream solution with wide user acceptance. It would service a geoscience knowledge system, meet users' requirements for linked handling of spatial, narrative and tabular information in the field or office, and fulfill the users' need for evaluated information and the suppliers' need to protect their reputation and intellectual property. The creation and implementation of new business, framework and surveying methods models is likely to take longer than the corresponding infrastructure developments, implying that pilot studies of new approaches are timely. Interacting aspects of the geological survey knowledge system, such as those discussed in this paper, rely on different areas of expertise. Geologists will be involved in extensive collaboration, investigation and experiment, as they climb the steep learning curve to advance with the cyberinfrastructure and meet fast-changing demands.

Acknowledgements

I am grateful to the editor and reviewers for their invaluable advice, and to colleagues, in particular John Laxton, Andy Howard, and Bill Hatton for their support. This paper is published by permission of the Director of the British Geological Survey (NERC).

References

- Brodaric, B., Gahegan, M., 2006. Representing geoscientific knowledge in cyberinfrastructure: some challenges, approaches and implementations, In: Sinha, A.K. (Ed.) *Geoinformatics: Data to Knowledge*, Geological Society of America Special Paper 397, pp. 1-20. doi: 10.1130/2006.2397(01).
- Foster, I., Kesselman, C. (Eds.), 2003. *The Grid: Blueprint for a New Computing Infrastructure*, 2nd edn., Morgan Kaufmann, San Francisco, 748 pp.
- Howard, A.S., Hatton, B., Reitsma, F., in press. Developing a geoscience knowledge framework for a national geological survey organisation. *Computers & Geosciences*, this volume.
- Loudon, T.V., Laxton, J.L., 2007. Steps toward Grid-based geological survey: suggestions for a systems framework of models, ontologies and workflows. *Geosphere* 3 (5), 319-336. doi: 10.1130/GES00095.
- Ludäscher, B., Lin, K., Bowers, S., Jaeger-Frank, E., Brodaric, B., Baru, C., 2006. Managing scientific data: from data integration to scientific workflows, In: Sinha, A.K. (Ed.) *Geoinformatics: Data to Knowledge*, Geological Society of America Special Paper 397, pp. 109-129. doi: 10.1130/2006.2397(08).
- National Research Council, 1993. *Solid-Earth Sciences and Society*. National Academy Press, Washington, DC, 346 pp.
- Raskin, R., 2006, Development of ontologies for earth system science, In: Sinha, A.K. (Ed.) *Geoinformatics: Data to Knowledge*, Geological Society of America Special Paper 397, pp. 195-199. doi: 10.1130/2006.2397(14).