

Suggestions for information architecture in systems geology

Contents

Suggestions for information architecture in systems geology	1
Contents.....	1
Overview of the architecture.....	3
What is information architecture?.....	3
Objectives of architectural design	3
Describing geology as a system	3
Viewing geology as a system.....	4
Representing geology as a system	4
An existing architecture	5
The role of geological surveying organisations.....	5
Structure of conventional Survey documents	5
Limitations of conventional documents	5
An information pool.....	7
Knowledge and communication	7
Sequences in text and images.....	7
Codifying information types	8
Coordinated information types	8
One modular pool, many narratives	9
Structure of linked data	9
Benefits of linked data in an information pool	10
A shared framework	11
The need for a shared framework	11
Objectives of the framework	11
Elements of the framework	11
Human working memory	12
The user interface	12
The user interface (graphic).....	13
A modular system	14
A problem shared.....	14

Modular writing	14
Types of information module	14
Modules and categories.....	15
Quantitative and qualitative modules	15
Geological map modules.....	15
Cross-references between modules	16
One source, many products.....	17
Single source documents	17
Reusable information.....	17
Information metadata.....	17
Mechanisms for future implementation	18
Why consider mechanisms now?	18
The purpose of markup languages.....	18
The development of markup languages	18
W3C, XML and HTML	19
XML related products.....	19
The DTD and Schema for XML	20
RDF and OWL	20
Specialised markup languages	20
DITA.....	21
Content management.....	21
The information ecosystem	23
Levels of formality.....	23
The evolution of information.....	24
Monitoring and updating.....	25
Bridges to earlier architectures	26
Two separate sources	26
Two sets of software.....	26
Links beyond the system.....	27
Conclusion.....	27
Summary of architectural needs.....	28
References	29
Glossary.....	30

Overview of the architecture

What is information architecture?

Information architecture is concerned with the form of representation, evaluation and evolution of information and its integration as an orderly, coherent structure that enables it to work for its designed purpose and to meet user needs.

Objectives of architectural design

Information technology can implement information structures to support various aspects of geology, including the technical documentation of systems geology¹ considered here. An architectural design aims to choose the most appropriate structures to achieve the intended benefits. The architecture also considers the procedures and mechanisms of documentation. It should conform to a consistent, evolving global view. It should answer such questions as:

1. What is the intended end-product of the documentation?
2. Is it appropriate for the likely requirements?
3. How can it best be achieved?
4. Will the result justify the effort?
5. How do we plan to get there, and why?
6. Can we design each step to take us closer to where we wish to be?

Describing geology as a system

Geology² is a science that describes a multidimensional system³ of intricately related components. Geological knowledge can be described, for example, in terms of:

- granularity (level of detail)
- spatial location
- spatial configuration⁴
- history through geological time
- material objects (rocks, minerals, fossils, etc.)
- processes (sedimentary, igneous, metamorphic, tectonic, palaeontological, etc.)
- geological provenance (viewing materials as the results of interactions of configurations of geological processes and pre-existing material)
- evidence and reasoning

¹ Systems geology: A view of geology re-based on the developing cyberinfrastructure and regarded as a system (a set of interacting parts that function as a whole) embedded in the wider knowledge system.

² Geology: Geology is defined in the AGI Glossary as: "The study of the planet Earth – the materials of which it is made, the processes that act on these materials, the products formed, and the history of the planet and its life forms since its origin. Geology considers the physical forces that act on the Earth, the chemistry of its constituent materials, and the biology of its past inhabitants as revealed by fossils...The knowledge thus obtained is placed in the service of man – to aid in the discovery of minerals and fuels of value in the Earth's crust, to identify geologically stable sites for major structures, and to provide foreknowledge of some of the dangers associated with the mobile force of a dynamic Earth." (Bates and Jackson, 1980).

³ System: A set of interacting parts that function as a whole. The systems approach involves study of linkages or interfaces between the component activities.

⁴ Configuration: The spatial arrangement, pattern, form and shape of objects, processes or their properties.

Viewing geology as a system

The systems approach to geology (outlined in BGS, 2011⁵, and explored in more detail in Loudon 2011⁶) represents information in such a way that the user can identify a particular topic, and follow its relationships and interactions. For example, the user might study the geological history of a particular area, and then zoom out to relate it to its broader context, examining the changing configurations of geological objects and processes through geological time. These should form a plausible evolutionary sequence, consistent with the present-day geology viewed at any scale. As another example, the user might examine a particular lithology (or the consequences of a particular process) as it appears in an area of immediate interest, compare it with occurrences in the surrounding areas, and look in detail at an observed sedimentary structure in the context of the processes and environment of its origin. Tracking such ideas through conventional reports is seldom straightforward. The user might also wish to investigate links to larger systems (such as Earth systems), and interactions with other systems (such as biological systems and economic systems). Systems documentation supports powerful methods for recording, disseminating and investigating multidimensional systems information of this kind.

Representing geology as a system

The [Limitations of conventional documents](#) 5 suggest that they are inadequate for recording and representing the complexity of interrelationships in systems geology. The cyberinfrastructure⁷ can support information architecture that is better suited to that task. The concepts of systems and their representation of information have been studied for many decades. The information architecture of shared knowledge systems is now a basis for coordinating essential information within fields such as: finance; design and maintenance of automotive and aerospace products; and software systems. Geological applications of systems techniques, less driven by necessity, have largely been in academic studies, with comprehensive applications emerging only in the last few years. The systems approach offers an integrating structure that can overcome constraints of conventional representation, and seems well suited to integrating representations of regional geology worldwide. The cyberinfrastructure has now advanced to the point where consideration of the information architecture of systems geology is appropriate. Geological survey organisations are well placed to take the next steps made possible by the advancing cyberinfrastructure. Initially, methods of systems geology can be linked to the application of powerful computer techniques for data collection, analysis and prediction. The underlying objective could be the development of a shared, authoritative, multidimensional systems model of the solid Earth. Some implications for the information architecture are considered here.

⁵ BGS, 2011. The basics of systems geology. <http://www.bgs.ac.uk/systemsgeology/basics.pdf>

⁶ Loudon, T.V. (2012). *A Scenario for Systems Geology*. BGS Research Report, RR/11/05.

www.bgs.ac.uk/systemsgeology/scenario.pdf

⁷ Cyberinfrastructure: An integrated assemblage of computing, information and communication facilities, deploying the combined capacity of multiple sites to provide a platform to underpin research and discovery, typically with broad access and end-to-end coordination.

An existing architecture

The role of geological surveying organisations

The systems approach to geology is seen here from a viewpoint within a Geological Survey. A role of a geological survey organisation⁸ is to provide an authoritative and systematic view, based on field survey and coordinated by shared standards, of what is known of the regional geology within its global context. The results are published as a comprehensive record providing consistent information by means of the conventional infrastructure of published maps and associated explanatory documents, and their digital equivalents and extensions, such as databases and spatial models. The aim is a more complete and accurate record of the nature and consequences of the world-wide operation of interacting geological objects and processes. It depends on a consistent and comprehensive approach, with a disciplined methodology applied across coordinated investigations.

Structure of conventional Survey documents

Examples of documents giving a comprehensive account of the geology of a region are the Regional Geology and Map Sheet Memoirs published by the British Geological Survey along with the associated geological maps. Many diverse topics are assembled as one coherent narrative. Major rock units are generally described in stratigraphical order, with an introductory overview. Separate chapters, particularly in the more detailed 'Memoir', may deal specifically with other topics, such as applied geology, geophysics, igneous activity, or structure. Inevitably, many different aspects of the geology are intertwined, as they all may play a part in the geology of even a small area. For example, a single paragraph in a regional summary may mention several rock types, their location and configuration, age, topographic expression, mode of formation, properties and composition, sedimentary structures, fossil and mineral content, structural deformation, and aspects of the geological reasoning. Each text document is a self-contained entity, published separately from the geological maps and other sources to which it may refer. Inherent in this structure are the various [Limitations of conventional documents 5](#).

Limitations of conventional documents

The architecture of conventional documents constrains both authors and readers:

1. The logistics of publishing paper documents limits the extent to which the text can be integrated with separately published maps and diagrams, or quantitative information from, say, downhole logs or seismic surveys, even where they refer to the same geology.
2. When ideas are revised, for example by a fossil find that alters the stratigraphical correlation, it is not easy to identify the wider consequences and amend all related records.
3. Revision and updating of a published document may occur only after many years or decades, and the reader of an earlier edition may be unaware of the changes.

⁸ 'A GSO [geological survey organisation] is a not-for-profit government organisation responsible for a range of tasks that generally include: geological surveying (mapping) of the nation, state, or province; conducting geological research to support economic development, public health, and environmental protection; distributing geoscience information; and advising government at various levels regarding water, mineral and energy resources, environmental issues, and earth hazards.' (Berg et al., 2011. Synopsis of current three-dimensional geological mapping and modeling in geological survey organisations. Illinois State Geological Survey Circular 578 (page 3). <http://library.isgs.uiuc.edu/Pubs/pdfs/circulars/c578.pdf>)

4. Limited and inconvenient access to cited documents deters the reader from switching to and from other sources to obtain more detail or see the wider context.
5. A reader primarily interested in one aspect, say, the structural development, cannot readily disentangle that thread from the other strands of geological description and interpretation, or follow the selected thread to related parts of the same or other documents.
6. A reader interested in, say, the provenances of flute casts world-wide, cannot readily find, extract and relate the detailed information that may exist in numerous potential sources.
7. More generally, even where relevant information exists in conventional documents, its interpretation may depend on its context in the document as a whole, and it may be difficult to extract in a suitable form for an unforeseen requirement.

A modular systems approach could help to overcome such constraints (see [Benefits of linked data in an information pool 10](#)).

An information pool

Knowledge and communication

Human brain cells are highly interconnected, enabling the brain to deal routinely with complicated multi-dimensional networks of systems⁹. They enable individuals to remember and think about a topic of interest as part of the interconnected body of their own knowledge. But the capacity of each individual brain is limited. Individuals therefore specialise, and to extend the scope of their knowledge they communicate with other specialists. Communications arise in diverse contexts and must be filtered by the recipients to extract what is useful to them. In general, the information structures used for storage in human memory differ from those used for communication (although some thought processes, such as recollection and reasoning, may be structured as a communication while remaining internal to the brain). Unlike the brain's multidimensional network, communication in written and spoken languages relies on a linear structure, one word following another in sequential narratives¹⁰ that have a beginning, middle and end. Graphical depictions, such as maps, charts and diagrams do a better (but severely limited) job of representing connections within configurations¹¹ in space and time. Mathematics can represent shared concepts and analytical methods, but much geological knowledge is based on intuitive recognition of observed pattern and background reasoning, which in this context has so far no adequate mathematical representation. In comparison with conventional procedures, the cyberinfrastructure¹² offers more flexible methods to store, manipulate, combine and communicate the various types of information. Like the human brain, it can structure and organise information in an architectural design that stores information as a network, from which it can extract and communicate sequences of information to meet a diversity of specific user requirements (see [One modular pool, many narratives 9](#)).

Sequences in text and images

The sequential aspect of narrative communication is generally associated with spoken language and with written text where the order of words, sentences, paragraphs, sections and chapters is an essential feature of a conventional document, leading from initial ideas through to a conclusion. However, other information types such as geological maps, spatial models, or time and stratigraphic diagrams, can also be a basis for narratives. The user might follow a sequence on a geological map (scanning by eye, pointing with a finger or drawing with a pencil), perhaps identifying the types of rock observed on a walk along the shore, or the sequence of strata exposed in a cliff. Or the user might follow pointers through a general geological time diagram to throw light on the sequence of events and their context that created a particular geological outcome. In three or four dimensions, one spatial model is the potential source of a multitude of different narratives, and as Google Earth demonstrates, an electronic map need have no boundaries. Mathematical analyses and simulations

⁹ System: A set of interacting parts that function as a whole. The systems approach involves study of linkages or interfaces between the component activities.

¹⁰ Narrative: A message, story or discourse relating sequences of, for example, real or imagined events, descriptive features, or reasoning.

¹¹ Configuration: The spatial arrangement, pattern, form and shape of objects, processes and their properties.

¹² Cyberinfrastructure: An integrated assemblage of computing, information and communication facilities, deploying the combined capacity of multiple sites to provide a framework to underpin research and discovery, typically with broad access and end-to-end coordination.

are based on algorithms, which by definition, are also sequences (of instructions). Narratives drawn from an information pool should be able to link and present sequences that integrate many types of information (such as text, maps, images, data and processes), each depicted in an appropriate form.

Codifying information types

Geologists know more about their subject than they can express. They can codify¹³ only some aspects of that knowledge to communicate it to others or to remind themselves of what they once knew. Geologists can codify their knowledge as:

- narrative text
- spatial information (maps and models)
- geological time information (stratigraphic charts, tables)
- algorithmic information (workflows, geological process models)
- database information (properties of objects and processes)
- text annotations
- images
- reference information (glossaries, stratigraphic codes)
- links to expert human knowledge

Most of these could be accessed on paper or on screen. However, codification implies an organised system, and the system is changing in response to new tools from the technologies of information, communication and computing. Consequently the procedures for representing this knowledge must also change. The various forms can be presented to the reader as [Coordinated information types 8](#) in an integrated view, and the multilevel information structure can be rationalised.

Coordinated information types

Different information types, such as text, GIS¹⁴-based maps, images, geological time diagrams, models, and databases, can be shown together as static or interactive windows on the same screen. This suggests that each module of information submitted to the system (and used in communication) should be in the most appropriate format (see [Types of information module 14](#)). For example, the geographical distribution of a particular rock type need not be described in the text if a linked map module can show the same information more effectively. The infrastructure can already overcome many constraints of conventional representation. It also opens the prospect of coordinating all information types, including automated mathematical analysis and simulation, within one integrated information pool. This could bring the overall representation and the forms of communication into closer harmony with the thought processes and perceptions in the geologist's mind. It should represent the interconnections and limitations of knowledge more adequately and support rapid and accurate communication across a wide community of users.

¹³ Codify: to create a representation or record of something in a form appropriate to the organised system of which it becomes a part.

¹⁴ Geographic Information System (GIS): An integrated system for the capture, storage, management, retrieval, analysis, manipulation and display of geographically referenced spatial data and its attributes.

One modular pool, many narratives

Conventionally, information is represented in a manner suitable for communication (such as a book, paper, map or dataset) and is published and stored in the same communicable form. However, the sequential nature of a narrative is better suited to communication than to recording and storing the information content. Technical documentation can be structured and stored, not as a collection of narratives, but as modules¹⁵ in an interconnected system from which numerous separate narratives can be created. Because the modules have multiple links, the reader can branch off from one thread of narrative to follow another. Systems applications in many other fields show that this is not only possible but also essential for flexible handling of diverse and complex relationships. Many supporting tools (mostly proprietary, but based on the same standards) have been developed in the wider systems context. In geology, there is a prospect of an information pool acting as a single comprehensive and internally consistent source of interlinked modules of many information types within an organised systems framework. Communication of this information to users would involve extracting modules sequentially from the pool, either by end-users selecting links to follow trails of their own making, or by intermediaries creating exchangeable documents. The intermediaries could edit modules and add transitional material, to provide more coherent and attractive presentations (from which, of course, readers could branch off to follow up points of specific interest to them).

Structure of linked data

Geographical space and geological time are fundamental to our understanding of geology. Geological maps and spatial models, together with the stratigraphical record, therefore reflect a basic structure for geological knowledge and its digital representation. [Representing geology as a system 4](#) extends that structure, and integrates text, graphical and mathematical representations. An aim of the architecture is to ensure that these various representations of knowledge are appropriately used, linked and coordinated in one coherent system. A benefit of the systems approach is more adequate management of the complex links relating all aspects of geology. The information can be organised as modules, each focused on a single topic, using database and connectivity techniques to interconnect related modules as a network, and enable users to follow their own trails through relevant information. For example, a stratigrapher might follow (and perhaps add to) sequences of information relating to various aspects of the Upper Silurian. A structural geologist might do the same with the Caledonian orogeny, a palaeontologist with the distribution of *Monograptus*, a tourist with the basic geomorphology, and so on. Modules, such as one describing the lithology of a particular formation, might reappear in any or all of these trails. One promising approach to storing the various forms of geological information is 'linked data'¹⁶, which brings considerable flexibility and encourages reusable public ontologies¹⁷ and semantic links between sources (Bizer et al., 2009¹⁸). A technical account of implementing an advanced linked data

¹⁵ Module: A subdivision of a system that can be combined with others in various ways to perform different functions, can be independently replaced or upgraded, or can be plugged into the system to extend its functionality.

¹⁶ Linked data: A style of publishing and interlinking structured data by looking up URI's on the Web. As opposed to seeing a fixed set of data sources, 'Linked Data' aims to create an unbound global data space. It links things in the world through their descriptions, by means of structured data on the Web.

¹⁷ Ontology: A formal representation describing concepts, entities and relationships in a domain of knowledge, typically providing a more detailed and rigorous machine-readable specification than a thesaurus or taxonomy.

¹⁸ Bizer, C., Heath, T., Berners-Lee, T., 2009. Linked data – the story so far. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 5(3), 1-22. DOI: 10.4018/ijswis.2009070101. <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>

application for storing and communicating information (on programmes broadcast by the BBC) is given by Sinclair, 2009¹⁹ and its related references.

Benefits of linked data in an information pool

Systems and linked data can work well together. The systems view can provide a rigorous and coherent approach to the science. The linked data approach can position the results in an appropriate multidisciplinary, multinational context. Many users with diverse requirements could extract or contribute information relevant to their own work, linked to its context, and selected and delivered according to the user's specification, including level of detail, level of expertise, form of presentation, topic, and receiving device such as a desktop workstation or mobile. The systems approach implemented in a linked data context offers a new perspective for authors and users, by making available authoritative geological information in an organised framework within a single global information space.

The potential benefits, overcoming some [Limitations of conventional documents 5](#), include:

- embedding systems geology (providing an authoritative, organised, coherent and flexible representation of a specific body of knowledge) in a global information system
- greater overall coherence and consistency and reduced repetition through rapid access to a wide range of linked information sources
- rapid and straightforward updating at the level of a module and its connections
- flexible selection and retrieval of material from the pool
- integrated records and displays of various information types in appropriate formats (text, maps, tables, diagrams, graphs, images, etc.)
- explicit links between the modules to enable each of a wide range of users with specific requirements to select their own chosen sequences of modules drawn from the same comprehensive information pool
- flexible presentation of selected sequences of modules in various formats, such as a book, a downloadable document, a paper or mobile field guide, a user-selected application (app), or a hybrid text and spatial website

To gain these benefits, [A shared framework 11](#) for the information architecture is needed to define and organise the modular formats and structures in which information is collected, evaluated and represented.

¹⁹ Sinclair, P., 2009. Linked Data on the BBC <http://www.slideshare.net/metade/linked-data-on-the-bbc>

A shared framework

The need for a shared framework

Licklider²⁰ (1968) pointed out that many difficulties in communicating ideas and opinions arise from mismatches of concepts and mental models between the communicators. In order to achieve a better match, he proposed that information technology should maintain, share and manipulate an explicit external model where many minds could work together, weaving their threads of thought into the web of a shared knowledge system. Thanks to the advancing cyberinfrastructure, this aspiration is becoming a realistic possibility. A shared information pool for systems geology could address this need with a framework²¹ that can improve mutual understanding by categorising the content of the geological subject matter in a widely acceptable form.

Objectives of the framework

A framework that provides a shared, explicit structure to connect geological thinking with its computer representations (a system metamodel) has the following objectives:

- to provide a multidimensional map of content and terminology to guide users and systems developers in depositing information, finding it, and handling and processing it
- to support the representation of fragments of information as modules that can be reused in many different contexts, by identifying their relationships to one another
- to locate and help to integrate ideas, concepts, workflows of investigation, and threads of reasoning that are shared among the various models of the geoscience knowledge system
- to establish well-defined interfaces that enable the business, investigation and infrastructure models to work together
- to simplify, rationalise and coordinate the interfaces to applications
- to strengthen the support for quantitative analysis by formalising the relationships among datasets

Elements of the framework

To meet these objectives, the framework attempts to build on stable and uncontested elements and relationships among the findings of geology. Loudon (2012²², p.117) tentatively proposed aspects of the framework referring to:

- the geographical location and configuration of geological objects and processes
- the resolution or level of detail in scale-space²³, at which they are considered or observed

²⁰ Licklider, J.C.R., 1968. The computer as a communications device <http://memex.org/licklider.pdf>

²¹ Framework: A logical structure and guidelines giving a broad overview for classifying and organizing complex information, within which detail can be added as required.

²² Loudon, T.V. (2012). *A Scenario for Systems Geology*. BGS Research Report, RR/11/05. www.bgs.ac.uk/systemsgeology

²³ Scale-space: In GIS, scale is the ratio of the distance between two points on a map or image to the corresponding distance on the ground. In the study of scale-space, however, scale may refer to the level of detail detected by eye or instrument across a range of scale from the finest detail discriminated, to the entire image or field of view. Scale-space theory regards this range as a multi-resolution continuum, zooming in or out to reduce or increase the amount of detail using a filtering process. It studies the range of scales over which object types exist and geological processes operate.

- the observed and interpreted properties, composition and relationships of the objects or processes
- the interaction of geological processes with pre-existing configurations of objects and the resulting events and historical changes through geological time

A modular system 14 requires a description (information metadata) that links each module to a geological framework that identifies its relationships with other modules. The framework must be designed as a coherent structure supporting the modular information architecture, and possibly as a means of assembling information as a solid Earth systems model²⁴.

Human working memory

The user interface to systems geology should follow familiar patterns of human thought processes. Their relevance to long-term storage and retrieval of geological information is considered in section 3.3 of Loudon and Laxton (2007)²⁵. This referred to the findings of brain scientists who have been able to determine the areas of the human brain in which various types of information are stored and processed, using scanning techniques. Different techniques were used by Baddeley (2003)²⁶, based on observations and clinical tests of normal patients and a wide range of patients with communication disorders. He reviewed models for working memory, which ‘involves the temporary storage and manipulation of information that is assumed to be necessary for a wider range of complex cognitive activities.’ This short-term memory has been linked to temporary electrical activity, as opposed to long-term memory that involves more durable changes in the nervous system. The temporary working memory may also serve as a workspace for other complex activities, such as reasoning, comprehension, and long-term learning. The point where human working memory interacts with the computer is [The user interface 12](#).

The user interface

Suggestions for a user interface matching the structure of Baddeley’s model of [Human working memory 12](#) are illustrated in [The user interface \(graphic\) 13](#). The upper half of the figure is based on Baddeley’s model of the human working memory. The phonological loop is concerned with verbal and acoustic information (mainly sequences of written or spoken words). The visuo-spatial sketch pad is its visual equivalent for understanding spatial input, and the central executive directs the attention of the brain to one or the other. The episodic buffer forms a temporary storage system binding together information from long-term memory and from the visuo-spatial and phonological systems. Together they pass information between the temporary working memory and longer-term memory.

The lower half of the figure proposes a structure for a user interface that aims to reflect that model. It is assumed that two linked displays are used in separate windows or partitions within a window, one principally for graphical display, the other largely for text. In fact, the nature and number of windows would depend on the display device and the complexity of the information, but that does

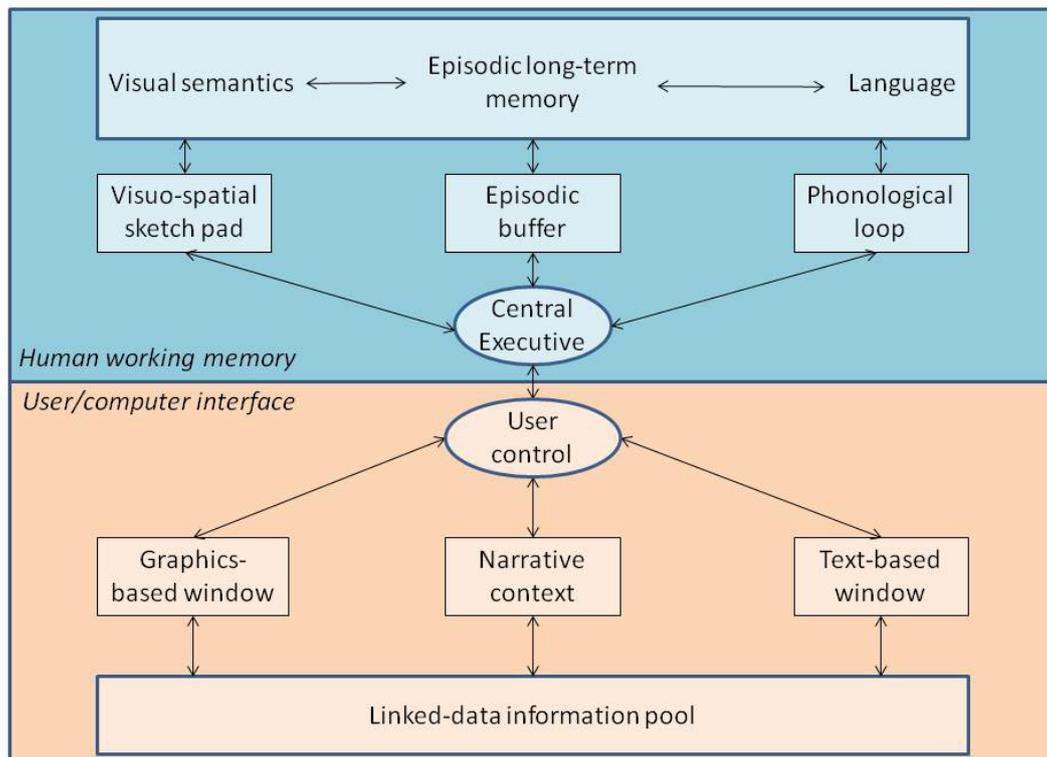
²⁴ Solid Earth systems model: A model of the systems of the solid Earth, organised within a framework or metamodel that depicts and clarifies the principal relationships among the findings of geology, providing a multidimensional map to locate and connect ideas, concepts, workflows of investigation and threads of reasoning.

²⁵ 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, pp.319-336; doi: 10:1130/GES00095.1 <http://nora.nerc.ac.uk/1084>

²⁶ Baddeley, A., 2003. Working memory and language: an overview. *Journal of Communication Disorders* **36**, 189-208.

not affect the underlying concept. The user would control the displays through the ‘user control’, such as selecting options from a menu with a keyboard or mouse. The information would be drawn from a linked-data pool in a sequence determined by a pre-existing trail or by a sequence of user decisions. In either case the context of the current position in the narrative would be held in a narrative-context buffer, which users could call up to determine the context of the current displays in the narrative sequence they had followed, and to back-track or follow a side-track if required. The narrative context could also clarify the route of the narrative and the available user options by locating the narrative in the more general framework of a solid Earth systems model.

The user interface (graphic)



A user interface and a model of working memory (based in part on Baddeley, 2003)

A modular system

A problem shared

A highly experienced geologist allegedly said that the most useful geologist is the one who has seen the most rocks. It has been suggested that the paper documentation for one type of aircraft weighed more than the aircraft to which it referred. True or not, the first statement suggests that the means of codifying geological information are inadequate for passing on geological knowledge and could benefit from improved methods of documentation. The second suggests that full documentation is labour-intensive and cumbersome, and it is therefore important to use the most effective methodologies. Of course, geological investigation is quite different from building an aircraft. One involves the information technology to capture, manage, store, preserve and deliver content and documents related to the organisation of objects and processes. But then, so does the other. The modular systems approach to technical documentation, which has proved effective in other fields, is likely also to be relevant in some aspects of geology.

Modular writing

According to Horn (2001)²⁷, technical writing should be modular, flexible and should fit into an appropriate structure. It should be precise enough that different writers working on separate modules could produce consistent results for plugging into the same document. Each module²⁸ (or information block, a paragraph or two in length) should deal with only one kind of functional meaning, leaving out anything that was not relevant to that specific function. Introductory phrases, transitional material relating the module to earlier and later paragraphs should be excluded, and definitions, examples and commentary should be in separate modules. The aim is to create a database of modules that can be used in many media and in different documents. Every module should have an informative label or subheading to make the organisation of the information more obvious, to help the reader to scan and skip, and to act as a transition from the previously selected module to the next one. Each module should relate to only one main point based on the purpose or function of the module. The same principles should be applied consistently to all modules.

Types of information module

The way in which sentences are related (stick together) was studied by Horn (2001) to define a taxonomy for types of information block or module. In the usual module the sentences are related by their functional concern with the same main subject. Other relationship patterns require the module to be structured differently. Decision tables represent if-then decisions; data tables make it readily possible to compare and contrast successive entries in rows and columns; argumentation maps clarify possibilities in disputed areas where conflicting opinions exist; topic maps identify modules related in various ways. The types of module differ for different kinds of technical knowledge. In geology, various other types of module are important, such as maps, logs, sections, time charts, datasets, and interactive modules with which the user could adjust the displayed content, or could contribute information to the system (see [Modules and categories 15](#)).

²⁷ Horn, R.E., 2001. What kinds of writing have a future?

<http://www.stanford.edu/~rhorn/a/recent/spchWhat%20KindsOfWrtng.pdf>

²⁸ Module: A subdivision of a system that can be combined with others in various ways to perform different functions, can be independently replaced or upgraded, or can be plugged into the system to extend its functionality.

Modules and categories

In other fields, text descriptions of system components typically refer to discrete objects and processes, to variations within them, and to their configurations, relationships and interactions. The systems in question generally refer to man-made entities, such as aerospace or automotive products, bank accounts, computer configurations or complex software systems. The man-made components are readily identifiable, with a clear process of manufacture, frequently by assembly from several identifiable sub-components. An example might be a wheel assembled from a steel casting, a rubber tyre, and bolts for attachment to the axle. Each sub-component is of a named type, identifying it as a member of a particular category (such as a specific type of bolt) with many individual, interchangeable items (as in a bucket of bolts of the same type), which, in the context of the application, could be regarded as identical. Analogous comments could be made about components in computer hardware or software systems. However, our knowledge of the natural systems of the solid Earth is based on observation and reasoning rather than manufacture. The objects, processes and configurations are not under human control, they show differences and discontinuities at every scale, and their full complexity is beyond human knowledge. Nevertheless to achieve some degree of understanding, we classify our observations into categories with more or less well-defined boundaries (such as stratigraphical, petrographical or palaeontological classes). These may be somewhat artificial, but can be named, described and widely shared, and reflect our knowledge of the geology. Likewise, a text account can be subdivided (as in man-made systems) into discrete modules dealing with separate topics – artefacts that make the text easier to organise, update, store, link and retrieve, and potentially easier to write, read and navigate.

Quantitative and qualitative modules

Quantitative measurements describe a continuum to any desired precision (though not necessarily accuracy). Qualitative statements, on the other hand, place the thing referred to into a category, the result of classification. For example, locations in space or time can be expressed quantitatively, whereas the geological formations that occupy that space are each given a (qualitative) name. A stratigraphical formation can be described in terms of its general or specific characteristics and variations in a particular area, each described in the relevant literature (such as a textbook on stratigraphy or a Survey memoir). Some properties of a formation, such as grain size, might be described quantitatively, others, such as being deposited by turbidity currents, might be stated qualitatively. The text description of, say, a geological formation may therefore relate to a qualitative concept, placed in a region within the quantitative three-dimensional continuum of space by maps and cross-sections, with colour-coding linking it to a stratigraphical key and quantitative scale of geological time. Qualitative and quantitative information are represented by different information types (such as text accounts, maps or graphs, databases) which can be shown simultaneously or sequentially in a narrative stream. Images and text are handled differently, both in the computer system and in human working memory, and it may sometimes be appropriate to view them together in separate windows (see [The user interface 12](#)).

Geological map modules

The conventional geological map is generally large, perhaps a metre across. The user can therefore stand back to view the patterns on the map as a whole, or can move closer to focus on the detail of a small area of the map while remaining aware of the surrounding content through peripheral vision. On a computer screen, on the other hand, the user of a geological map or spatial model sees only a

small area at a time. However, the user can scan across to other parts of the map, zoom in to see more detail or zoom out to see the wider context. Potentially, the spatial model provides the means to move smoothly through scale-space²⁹, as opposed to the fixed scales of conventional maps. The model also has the potential to provide mechanisms for users to select the topics to be shown on the displayed map, and to view the three dimensional geology at any level and cross-sections in any orientation. This, however, will require an appropriate mathematical representation of the geological information, as opposed to the geologists' hand-drawn interpretation.

Cross-references between modules

Inherent in the modular representation of information is the need for modules to reference one another, despite differences in their information types and scope. Thus, a text module describing a small area in detail might refer to a more general text module describing a wider area. However, it might also refer to a graphic module that has to be created on the fly by extracting a small portion from a continuous map or spatial model. As different information types (such as maps and text) can be displayed at the same time, possibly in separate windows, the user must be able to move freely from one type to another, and the modules must be able to cross-reference one another in their own ways. The cross-reference from a discrete text module to a region within a continuous spatial model would have to include the boundaries of the relevant area in an appropriate form for extraction through a GIS-based³⁰ interface. If user input were needed, it should provide the means for the user to select the area and content from an appropriate script. A reference from, say, a text module might also carry instructions to highlight appropriate parts of a map or other referenced module. Conversely, cross-reference from a spatial model to text modules requires GIS features to show the point or area referred to by the text module, with a clickable link as demonstrated, for example, by the BGS GeoIndex³¹. The flexibility of the systems approach encourages quantification of information where possible, thus giving access to powerful statistical and other mathematical models. Again, flexibility in referencing between diverse information types is required. It is likely that, say, a reference from a text module to a database would require the means to identify and extract the required data by criteria such as 'equal to', 'greater than' and 'less than'. Again, the parameters might be defined by the text module, or might be selected from a script by the user.

²⁹ Scale-space: In GIS, scale is the ratio of the distance between two points on a map or image to the corresponding distance on the ground. In the study of scale-space, however, scale may refer to the level of detail detected by eye or instrument across a range of scale from the finest detail discriminated, to the entire image or field of view. Scale-space theory regards this range as a multi-resolution continuum, zooming in or out to reduce or increase the amount of detail using a filtering process. It studies the range of scales over which object types exist and geological processes operate.

³⁰ Geographic Information System (GIS): An integrated system for the capture, storage, management, retrieval, analysis, manipulation and display of geographically referenced spatial data and its attributes.

³¹ GeoIndex: Search BGS datasets using GeoIndex; a map-based index of a wealth of information that we have collected or have obtained from other sources. <http://www.bgs.ac.uk/GeoIndex/home.html>

One source, many products

Single source documents

Single sourcing refers to recording information once and delivering it in various formats. For example, geologists might each draft a separate chapter of a regional guide, sharing the same style guide and word processing software. The results might be delivered to a publisher who converted it to different software to be edited, revised, assembled and delivered in book form. The same material might also be delivered as an ebook. Sections of it might also be made separately available in a different layout on a web site. The same content is a single source that passes through various conversions from one software system to another, and is delivered in various manifestations. A narrative resulting from selecting a sequence of modules from a linked data information pool might be made available in many forms, such as a book, report, or website presentation. It might be accessed through a wide range of devices from desk-top computer to smart phone. The modules of systems geology should therefore be single source documents, and also [Reusable information 17](#).

Reusable information

Reusable information, in this context, refers to information that can be reused in many different contexts without rewriting. The author of reusable information cannot make assumptions about the audience for the products in which the information may be used, or for the contexts in which the information may be placed. This requires significant changes to the architectural design and representation of the information. Typically, reusable information would be written in modular form, with each module making sense as stand-alone information, and maintained and updated as a separate item. In order to identify its relevance to potential applications, each module in an established system (sometimes referred to as a 'topic' in this context) requires comprehensive information metadata. Modules in systems geology would normally be [Single source documents 17](#), providing [Reusable information 17](#) within [A shared framework 11](#) defined by [Information metadata 17](#).

Information metadata

Information metadata in this context is a higher-level description of the information contained in the module (see [RDF and OWL 20](#)). It would record such aspects as:

1. The information type (see [Types of information module 14](#)) identifying the software required for its display or processing
2. The provenance (author, date, editor, publisher, and possibly project, workflow, and other contributors)
3. The geological relevance (position in [A shared framework 11](#))
4. The relationships to other modules (such as links to associated modules of different information type, or position in a hierarchy such as level of detail or [Levels of formality 23](#))
5. The references or links to and from other modules
6. The references to published documents or other entities outside the system
7. Links to archive copies of all earlier approved versions, as it may be necessary to look back to earlier ideas in their wider context to understand the reasoning behind previous interpretations

Mechanisms for future implementation

Why consider mechanisms now?

A full implementation of new information architecture for systems geology would clearly be a huge task, involving a wide range of technical and geological experts. An underlying objective of systems geology is to develop a unifying environment where the shared significance and potential of many separate viewpoints and sources of information come together to improve our geological understanding. This implies wide geographical and interdisciplinary collaboration to discourage the emergence of conflicting, incompatible systems. Essential precursors to implementation include content and cost-benefit analyses, scientific justification, feasibility and pilot studies, and much more. Before that is considered, however, there should be a clear view of the intended long-term objectives and their implications, a perception of how they can be achieved, and confidence that the tools for making the transformation will be available. That is the aim of a scenario – a thought experiment such as this one – where ideas can be developed, explained, assessed, modified, and discarded or adopted at minimal cost. In [Mechanisms for future implementation 18](#), therefore, some required mechanisms are outlined that might also prove helpful to those thinking only of shorter-term adjustments to their ways of working.

The purpose of markup languages

A markup language provides a set of rules for encoding documents in a machine readable form. It defines a set of annotations for computer instructions that can be inserted into the body of a text document. The annotations are placed within the text, usually between angle brackets that enable the software to recognise and act upon them. They are normally omitted from the display for the human reader but may be included for proof-reading, and are therefore designed for easy human recognition. The user of a word processor might select, say, 'section break' from a menu, and the word processing software would automatically insert the appropriate annotation in the text at that point. Thus <section> might be inserted to indicate to the computer a point where a new section of the text begins and </section> might indicate the end of that section. A program responsible for rendering the document in print uses the annotations to control the layout, according to requirements specified by default or by the user. The markup language thus separates the form of presentation from the form of recording. The annotations are not restricted to specifying the form of display. For example, the URL links in the hypertext markup language (HTML) of a Web page can locate and display other documents.

The development of markup languages

Automation of the processes of publication, maintenance, and exchange of business documents brought the need to describe the structure and other attributes of a document (independently of how it should be processed or delivered) and led to the development of the Standard Generalized Markup Language (SGML). SGML became an ISO standard in 1986 – comprehensive, complex and difficult to use. However, it was the progenitor of a range of markup languages that were easier to use and understand. The most familiar (from its widespread use on the Web) is HTML, which has the considerable merit of simplicity, but is focused on the display of information. Communities requiring markup to meet their own individual needs had to move beyond the simple tag set of HTML, while conforming to a shared set of rules. This led to XML, a simplified subset of SGML suitable for web

applications. It provides the means for integrating heterogeneous applications and exchanging information: 'an open adaptable format for encoding data that could be exchanged using a wide range of mechanisms and imported into a variety of applications' (Gollner, 2010³²).

W3C, XML and HTML

The World Wide Web Consortium ([W3C](#)) develops Web standards through agreed specifications ('Recommendations') that describe communication protocols and other building blocks of the Web. Their working groups coordinate their activities through a wide membership of organisations, which include Web users and developers, and commercial, research, academic, government, and standards bodies. The W3C Recommendations include the extensible markup language (XML). Derived from SGML (see [The development of markup languages 18](#)), [XML](#) is a software- and hardware-independent markup language for structuring, transporting and storing information. The advantages of XML include making it easier to:

- share data³³ (in the computer science sense) from different sources
- upgrade to new hardware, operating systems, applications or browsers
- make data available to a wide range of devices
- link to languages (such as [RDF and OWL 20](#)) readable by computer applications.

On the other hand, the layout and display of information for Web applications, such as information stored and transported in XML, can be handled by [HTML](#). XHTML is a more rigorous version of HTML that is compatible with XML and should be rendered correctly even on handheld devices with limited computing power, such as mobile phones.

XML related products

A sitemap for some [tutorials](#) for W3C standards links to a range of [XML](#) related products; for example:

- for querying information ([SQL](#) , [Xquery](#))
- scripting languages to add functionality to a Web page ([Javascript](#))
- external style sheets that define how to display the content ([CSS](#))
- a standard way for accessing and manipulating XML documents ([DOM](#))
- protocols for accessing a Web service ([SOAP](#))
- handling multimedia (anything you can see or hear)
- making dynamic and interactive Web pages ([PHP](#))
- transforming XML into HTML or another language ([XSLT](#))
- an XML syntax for interchange of topic maps³⁴ ([XTM](#))

³² Gollner, J., The emergence of intelligent content <http://jgollner.typepad.com/files/the-emergence-of-intelligent-content-jgollner-jan-2009.pdf>

³³ Data: A collection of observations, measurements or other information about a set of variables, generally in the form of numbers, words, or images. In the computer science sense, data is information suitable for use by a computer, distinct from the program (a set of instructions to perform a task).

³⁴ TopicMaps.Org Authoring Group, 2001. XML Topic Maps (XTM) 1.0: TopicMaps.Org. <http://www.topicmaps.org/xtm/index.html>

[Trippe \(2008\)](#)³⁵ reviews ‘the challenges faced in managing information in robust systems that enable reuse of XML-encoded data objects and support their efficient processing. This is accomplished through the use of XML databases and open standards based on the eXtensible Markup Language (XML) produced by the World Wide Web Consortium.’ He points out that XML has a hierarchical structure that can be used to model virtually any type of data. It enables the user to take advantage of the many W3C standards to provide efficient and precise access, query, storage, and processing capabilities not found in traditional database technology.

The DTD and Schema for XML

To avoid the complexity of a fully generalised markup language, communities of users can specify markup procedures for their own documents using a DTD. An XML document type definition ([DTD](#)) is a file that defines the allowable markup and markup rules followed in an XML document. The file may be internal to the document or held in an external file that can be referred to by many documents. However, the DTDs have now been largely superseded by the more powerful XML schemas. They are themselves written in XML, and are the means for the sender to describe data in a way that the receiver can understand. A document can reference several schemas. A schema can be extended by future additions, and can be reused in other schemas. It supports data types, making it easier to describe and validate the content and to work with data from a database. The development of a complex DTD or schema is laborious and time-consuming, and may require changes as new requirements arise. Many users therefore use an existing general schema, without the benefits of specific markup (but see also [Specialised markup languages 20](#) and [DITA 21](#)).

RDF and OWL

The Resource Description Framework ([RDF](#)) is a W3C recommendation for describing Web resources in a way that can be read and understood by computer applications. RDF documents are written in XML, and can therefore be readily exchanged between computers, operating systems and application languages. The RDF description for a Web page might, for example, contain information regarding its content, author, version and date. It is part of the W3C’s Semantic Web Vision, where the information has exact meaning, and can be understood, processed, and integrated with other files by computer. RDF statements follow the form of subject, predicate, object sequence described in the semantic Web, and identify things by means of identifiers (URIs³⁶). The RDF Schema provides the framework to describe application-specific classes and properties. RDF can, for example, link to [Dublin Core](#) metadata, a widely used set of predefined properties for describing documents (potentially including systems geology modules). The Web Ontology Language ([OWL](#)) is a W3C recommendation, written in XML to provide a larger vocabulary and stronger syntax than RDF with greater machine interpretability.

Specialised markup languages

A wide range of XML-based markup languages has been developed for specialised content (see [The Linux Foundation list](#)). Those of particular interest in the present context include the following:

³⁵ Trippe, B., 2008. Using XML and databases: W3C standards in practice. The Gilbane Group, White Paper.

<http://gilbane.com/whitepapers/EMC/EMC-XML-whitepaper-4.08.pdf>

³⁶ Uniform Resource Identifier (URI): A string of characters giving an address on the Web that identifies a physical or abstract resource (such as a thing, document, image, or service). A Uniform Resource Locator (URL) is a URI that provides the means of locating the resource.

1. [Office Open XML](#) file formats were introduced in Microsoft Office 2007 products (Word, Excel and Powerpoint) so that documents prepared using these products could be represented in an XML-based format.
2. The [Geography Markup Language](#) (GML) was developed by the OpenGIS Consortium as an XML encoding for the transport and storage of geographic information, including both the spatial and non-spatial properties of geographic features. It is concerned with the properties and geometry of objects, as opposed to the symbols and graphic representation of them on a map.
3. The [Geoscience Markup Language](#) (GeoSciML) is a GML Application Schema that can be used to transfer information about geology, with an emphasis on the "interpreted geology" that is conventionally portrayed on geological maps.
4. The Keyhole Markup Language ([KML](#)) was developed for use with Google Earth and is now a standard of the [Open Geospatial Consortium](#). It is a file format for display of geographic data in an Earth browser such as Google Earth. It is widely used for display of Earth science data, for example by the US Geological Survey (see Ballagh et al., 2011³⁷).
5. The [Mathematical Markup Language](#) (MathML) is an application of XML for describing mathematical notations and capturing their structure and content. The goal of MathML is to enable mathematics to be served, received, and processed on the World Wide Web, just as HTML has enabled this functionality for text.
6. The [Systems Biology Markup Language](#) (SBML) is designed to enable computational models of biological systems to be shared and published in a form that other researchers can use, possibly in a different software environment.

DITA

The Darwin Information Typing Language ([DITA](#)) is an [OASIS](#) standard for creating and publishing modular information as XML files. An appropriate application is to structured authoring, where a DTD or schema can capture the rules for organising the information. The content can be single-sourced, reusable, (see [One source, many products 17](#)) and appropriately organised as modules, referred to in DITA as 'topics'. The topic is the basic architectural unit, and each topic belongs to a pre-defined category ('topic type'), such as a concept, task or reference. The author can describe relationships between topics. Each topic is written once but can be linked to other topics in various ways. Rather than creating a DTD or schema from scratch, DITA provides DTD's for a standard set of basic topic types, any of which can be extended ('specialized') to meet more specific needs. 'This concept allows writing teams to change or adapt the DTD to their needs over time, without requiring expensive, complicated rewrites of the entire DTD.' ([Fisher, 2003](#)). The official DITA [website](#) provides up-to-date information on developments.

Content management

Content has been defined as the persistent expression in physical form of intended meaning in information exchange, as opposed to its format or method of delivery. Content management is concerned with the technologies for handling the content and the resulting products, such as formal publications, documents, manuals, informal and formal records, field notes, specimen descriptions, maps, diagrams, photographs, databases, images, forms, emails, audio, and video. 'Enterprise

³⁷ Ballagh, L.M., et al., 2011. Representing scientific data sets in KML: Methods and challenges. *Computers & Geosciences*, **37**, 57-64. doi: 10.1016/j.cageo.2010.05.004

content management' is concerned with managing the content in a large organisation as one integrated activity, where it can work alongside the management of other formal activities, such as resource planning or customer relationships. 'Component content management' is concerned with management of content 'at a granular level' (Trippe, 2005³⁸), where content is divided into smaller modules or components. The aim is to gain flexibility in creating [Single source documents 17](#) and [Reusable information 17](#). 'Each component can be separately edited and re-used, and workflow processes enforced. Content components can have their own lifecycles and properties – version, owner and approval – that support fine-grained reuse and the ability to track such usage.' Such an approach is clearly relevant to systems geology, where the modular structure can be predefined across all information types. It would, however, be difficult to apply to much existing geological documentation where many topics are inextricably intertwined and major rewriting would therefore be required.

³⁸ Trippe, B., 2005. Component content management in practice: meeting the demands of the most complex content applications. The Gilbane Report: White Paper. http://gilbane.com/whitepapers/X-Hive/Xhive_Gilbane_Whitepaper_CCMS6-final.pdf

The information ecosystem

Levels of formality

The structure of the present system for communicating geological information has several levels of formality, from widely agreed and evaluated authoritative documents to individual suggestions or findings of limited scope that may make sense only with further explanation by the originator or the curator. Geological survey organisations might consider, discuss, explore and eventually develop and maintain a structure to implement shared information architecture for collaborative systems geology. The systems structure would follow closely the existing [Levels of formality 23](#) but would place greater emphasis on global and interdisciplinary coordination and on working in parallel with mainstream developments in the cyber-infrastructure.

Examples of levels of formality in a geological surveying organisation might include:

1. The scientific standards to which information content should conform, generally involving international and cross-disciplinary collaboration. These include standard classifications, ontologies, codes, glossaries, etc. for stratigraphy, petrology, mineralogy, cartography, etc., evaluated by wide consultation and authorised by standard setting bodies; and styles, rules and conventions approved and required by in-house management.
2. The consistent, systematic and authoritative account of the relevant regional geology, and related surveys of geophysical, geochemical and other properties. The input is fully reviewed and evaluated before acceptance, and thereafter kept up to date and available; maintained in-house, generally evaluated by experts following house rules, and authorised and published by Surveys in the form of maps and documents such as map explanations.
3. Links to published scientific papers, typically dealing with some specific aspect of the regional geology, following defined conventions, and evaluated and authorised by the journal editor assisted by expert reviewers. The content should be accessible in a research archive (such as NORA³⁹), and hyperlinked as appropriate to and from the systems documentation, recording specialist findings that are relevant to the system.
4. Supplementary material such as archived records of detailed field notes, descriptions and analyses of samples and specimens, datasets, photographs, investigational workflows, borehole logs, specimens, cores and samples. Hyperlinked to the system and accessed as appropriate, if need be with expert guidance.
5. Drafts of documents in preparation or under review.
6. Correspondence, discussion, informal communications, presentations, etc.
7. Discussion and informal material such as alternative opinions and new evidence not fully taken into account, but maintained and monitored.

The confidence that can be placed in information depends on assessment of the reputation and quality of the organisation providing the information, in turn reflecting the rigour of the procedures in place to monitor its quality. The confidence obviously differs at each level of formality and each stage in [The information life cycle 24](#).

³⁹ NORA, 2010. The NERC Open Research Archive. <http://nora.nerc.ac.uk>

The information life cycle

The life cycle of information involves its development and modification, with increasing [Levels of formality](#), from collection and recording in the field and office through to the stages of editing, review, evaluation, revision, approval, incorporation in the system, archiving, delivery to users, revision or disposal. The workflow through this life cycle is designed to control access to appropriate individuals at each stage, to ensure that appropriate software is available at every stage, and ensure that information can flow freely across the interfaces between stages. A disciplined approach is needed to monitor and assess quality, and to structure correctly the complex interconnections. In scientific papers, quality is maintained by editorial judgment and peer review; and the information evolves largely by citation in subsequent papers that reinforce or modify the scientific content. On the other hand, documentation published by Geological Surveys or similar organisations is generally reviewed in-house (where expertise should be available to maintain consistent quality throughout the life cycle) and evolves by periodic revision. The modular approach of systems geology is well suited to Survey publications. Published material is normally archived at the level of the minimum revisable unit⁴⁰, which in systems geology would correspond to the module and its links; in conventional publication it is a published document, such as a paper, book or map.

The evolution of information

“Content⁴¹ populates an ecosystem where people receive, internalize, modify, use, create and share that content. In this way, content evolves.” (Gollner⁴², 2009). Different species of content evolve at different rates, not unrelated to the [Levels of formality 23](#). According to Kuhn⁴³ (1962), normal science rests on a paradigm – the common beliefs and agreements shared between scientists about how problems should be understood and addressed. Paradigm change is a most infrequent event resulting from a breakdown of the explanations of normal science, and leads to major upheaval and reconsideration of earlier work. It makes sense, therefore, to standardise undisputed elements of geology in a systems framework. The more transient details of local geology, however, may be uncertain interpretations of inadequate data subject to repeated revision. The modular approach speeds the movement of this transient information through its life cycle. Information can readily be amended and knock-on effects remedied by designing the modules as minimum revisable units, and recording the content separately from its connections. Alternative explanations and uncertainty are easier to handle in digital spatial models than in the conventional geological map. Nevertheless, the aim of providing a coherent regional view of complex global systems requires a wide range of collaborating experts. It may be most readily achieved and documented by a long-standing surveying organisation. On the other hand, the publication of scientific papers provides the diversity, selection and heredity that enable geological ideas to evolve through subsequent quotation and citation. Informal communication and communal authoring, such as blogs and Wikipedia, are developing as

⁴⁰ Minimum revisable unit: A self-contained subset of information in a documentation system: a component of the system designed to be revised as necessary without endangering the integrity of the system as a whole.

⁴¹ Content: the persistent expression in physical form of intended meaning in information exchange, as opposed to its format or method of delivery.

⁴² Gollner, J., 2009. Content fusion: There’s a piece of data lodged in my document. Intelligent Content 2009, Palm Springs, CA. www.slideshare.net/jgollner/content-fusion-or-there-a-piece-of-data-lodged-in-my-document

⁴³ Kuhn, T.S., 1962. The structure of scientific revolutions. The University of Chicago Press, Chicago, 172pp.

another layer of information, with consequences that have not yet been fully explored (Procter et al.⁴⁴, 2010).

Monitoring and updating

Systems documentation requires constant and consistent monitoring, correcting and updating to ensure that recent, accurate, coherent and consistent information is always available. These responsibilities are more easily discharged within a specialist organisation. As with scientific journals, it is important to maintain an archival record of superseded systems information, as reference to earlier records may be essential for understanding the development of ideas, their influence on past decisions and their current validity. This process is readily automated in systems documentation, as for example, in [MediaWiki software](#). Systems information can also be linked to informal and external communications, for example from casual users in the field, possibly equipped with smart phones including camera, communication, and location finding devices. They might wish to record their own observations as well as viewing existing information. For example, they might observe features that seemed to conflict with the official view or notice new exposures during road construction. While the official view must be maintained within firewalls and paywalls to maintain a consistent, authorised archive, it could also be linked to informal records in separate discussion areas: outside the firewalls, open to all, and monitored to update the authoritative view where appropriate. The framework of systems geology could provide links from the current authoritative view to alternative evolving interpretations, thus highlighting areas of dispute or in need of revision.

⁴⁴ Procter, R., Williams, R., Stewart, J., Poschen, M., Snee, H., Voss, A., Asgari-Terghi, M., 2010. Adoption and use of Web 2.0 in scholarly communications. *Phil. Trans. R. Soc. A* **368**, 4039-4056.

Bridges to earlier architectures

Two separate sources

Systems geology is based on a structural framework of geological concepts, designed for storage and analysis. It aims to populate this framework with modules that can be clearly positioned within its structure. Storage and analysis of the information are separated from its communication (see [Knowledge and communication 7](#)). The systems approach seems particularly appropriate for the flexible representation of methodical regional geology. But even in Surveys, almost all the relevant information is at present outside the proposed architecture, much of it in geologists' minds or in conventional paper form. Considerable skill and expert knowledge are needed to convert knowledge from existing text documents into a modular form fully compatible with systems architecture. Furthermore, the contents of existing publications are expressions of individual views at a particular time. They are archived as an inviolable record, which can be cited in other documents but where fragments of information can mislead if taken out of context. Communication is paramount in conventional scientific publications. Such publications are likely to start from observations and ideas, and weave together many concepts in sequence, to arrive at conclusions that extend or clarify existing knowledge. They are well suited for many purposes, and likely to remain in use for the foreseeable future. The need therefore arises of building bridges between systems geology and traditional publications: bridges that help users to integrate both sources of information, and that can support migration of information from existing sources into the systems framework.

Two sets of software

Conventional documentation and systems geology can be recognised as [Two separate sources 26](#) of information. Sharing their metadata would be an important step in bridge building. But two separate software systems have emerged to organise their content: topic maps and [RDF and OWL 20](#). They came from different communities. Topic maps⁴⁵ arose from work on managing indexes to documentation. The Resource Description Framework (RDF) evolved from assigning descriptive metadata to collections of web pages and websites, extended (with OWL) by the artificial intelligence community.

Classification of documents follows the initial writing and is traditionally undertaken by librarians, archivists, and authors who prepare indexes to their books. Topic maps (along with tools for ontological searching⁴⁶ and concept mapping⁴⁷) developed from this classification process, arising from work on managing indexes to documentation. The word 'topic' is used in this context to refer to a 'thing' about which something can be asserted. It is the object or node in the topic map that refers to a single subject. Topic maps extend the concept of electronic indexes to include the capabilities of thesauri, taxonomies, and faceted classification⁴⁸. They overcome the distinction between individual documents, addressing the need to span multiple documents and cover large pools of information. Their purpose is to describe knowledge structures and associate them with information resources. Topic maps became an ISO standard in 1999. Subsequent work (Garshol,

⁴⁵ Ahmed, K. Moore, G., 2005. An introduction to topic maps. <http://msdn.microsoft.com/en-us/library/aa480048.aspx>

⁴⁶ Calais: connect everything. 2012. <http://www.openalais.com/>

⁴⁷ Concept map: a diagram showing the relationships among concepts. http://en.wikipedia.org/wiki/Concept_map

⁴⁸ Pepper, S., 2002. The TAO of topic maps. <http://www.ontopia.net/topicmaps/materials/tao.html>

2008⁴⁹) provides methods of converting information between topic maps and RDF (the W3C standard). It remains to be seen whether topic maps will have a future place in systems supported by RDF and OWL.

Links beyond the system

Geological maps are routinely digitised and can be made available with GIS⁵⁰ support as a more comprehensive spatial model. Field and laboratory data, including digital photographs, are routinely stored in a database management system. Geological maps and databases readily conform to the systems approach. Procedures for collecting field data are evolving and could usefully take into account longer-term views on architectural design, and vice versa. Technology now offers the flexibility of recording, analysing and displaying many information types side-by-side, and some artificial barriers imposed by earlier technology will become irrelevant and unhelpful. Yet many geological organisations continue to have separate groups specialising in such fields as website development, database management, cartography and GIS, and editorial and publications. Systems geology suggests that, despite information being stored and communicated in separate structures, information content should be fully integrated, throughout [The information life cycle 24](#) and across all information types. For example, classifications of text documents could be matched with the ontologies of the systems framework. Although Surveys cannot alter the content of existing text publications, they can and do reformat the content as digital documents, by scanning, optical character recognition, and digital archiving. Access to the diverse information might also be improved, for example, by topic analysis with widely available software and coordinated within an appropriate framework (see [The need for a shared framework 11](#), [Two sets of software 26](#)). Categorisations of existing text documents might be aligned with the chosen systems framework, to simplify cross-referencing between systems geology and the geological literature. The combination of digitisation and semantic linking provides an accurate search tool and a means of coordinating different information types and representations. It adds longer-term value, and could benefit greatly from integration with the flexibility and clear structure of the systems approach.

Conclusion

The information architecture must be sufficiently flexible to accommodate information in conventional form and in intermediate formats, while encouraging convergence on more advanced systems. It should also aim to converge with the longer-term objectives, standards, and conventions of the evolving cyberinfrastructure and widely used software systems, in order to ensure a future migration path and reduce future back-tracking. It must be coordinated globally in both the geographic and the disciplinary sense. However, just as global implementation of standardised stratigraphic codes and geological maps took many decades to achieve, so also the development of mature global standards for systems geology, ontologies and information architecture will inevitably take considerable time and effort.

⁴⁹ Garshol, L.M., 2008. Living with topic maps and RDF. <http://www.ontopia.net/topicmaps/materials/tmrd.html>

⁵⁰ Geographic Information System (GIS): An integrated system for the capture, storage, management, retrieval, analysis, manipulation and display of geographically referenced spatial data and its attributes.

Summary of architectural needs

1. The architecture should ensure that systems geology is represented in a form appropriate for international and interdisciplinary collaboration and access, and implemented using shared mainstream products that will support future migration.
2. The architecture should support the operation of an information pool of modular systems that integrate information of many information types, with flexible means of tracking and extracting narratives for communication in appropriate formats.
3. The system should be structured within a stable framework of widely agreed, uncontested concepts and standards supporting a coordinated world view of a solid Earth systems model.
4. The user interface should handle the extraction of segments from continuous records in databases and representations of space and geological time (such as maps and cross-sections) for display along with their corresponding text and other relevant modules.
5. The implementation should be based on, and updated in parallel with, familiar mainstream technologies, and embedded in the global development of linked data.
6. The information architecture for systems geology should meet global standards for information interchange, and should ensure that geological representations are compatible with their wider context, including procedures for citation and linking.
7. The system should readily link to and from conventional representations and procedures.
8. The system should provide a migration path for reworking conventional representations to make them more readily accessible and useful in the systems context.
9. The architectural design should support the information through its complete life cycle from origin to obsolescence.
10. Mechanisms must be in place to ensure that geological systems information is of high quality, acceptable, relevant, up-to-date, coherent, consistent and readily accessible.
11. The information ecosystem should ensure that users can readily discover and possibly influence an informed assessment of the significance and reliability of the content.

References

- Ahmed, K. Moore, G, 2005. An introduction to topic maps. <http://msdn.microsoft.com/en-us/library/aa480048.aspx>
- Baddeley, A., 2003. Working memory and language: an overview. *Journal of Communication Disorders* **36**, 189-208. doi:10.1016/S0021-9924(03)00019-4
- Ballagh, L.M., et al., 2011. Representing scientific data sets in KML: Methods and challenges. *Computers & Geosciences*, **37**, 57-64. doi: 10.1016/j.cageo.2010.05.004
- Berg et al., 2011. Synopsis of current three-dimensional geological mapping and modeling in geological survey organisations. Illinois State Geological Survey Circular 578 (page 3). <http://library.isgs.uiuc.edu/Pubs/pdfs/circulars/c578.pdf>
- Bizer, C., Heath, T., Berners-Lee, T., 2009. Linked data – the story so far. *International Journal on Semantic Web and Information Systems (IJSWIS)*, **5**(3), 1-22. DOI: 10.4018/ijswis.2009070101. <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>
- Garshol, L.M., 2008. Living with topic maps and RDF. <http://www.ontopia.net/topicmaps/materials/tmrdf.html>
- GeoIndex: Search BGS datasets using GeoIndex; a map-based index of a wealth of information that we have collected or have obtained from other sources. <http://www.bgs.ac.uk/GeoIndex/home.html>
- Gollner, J., 2009. Content fusion: There's a piece of data lodged in my document. Intelligent Content 2009, Palm Springs, CA. www.slideshare.net/jgollner/content-fusion-or-there-a-piece-of-data-lodged-in-my-document
- Gollner, J., The emergence of intelligent content <http://jgollner.typepad.com/files/the-emergence-of-intelligent-content-jgollner-jan-2009.pdf>
- Horn, R.E., 2001. What kinds of writing have a future? <http://www.stanford.edu/~rhorn/a/recent/spchWhat%20KindsOfWrtng.pdf>
- Kuhn, T.S., 1962. The structure of scientific revolutions. The University of Chicago Press, Chicago, 172pp.
- Licklider, J.C.R., 1968. The computer as a communications device <http://memex.org/licklider.pdf> 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, pp.319-336; doi: 10.1130/GES00095.1 <http://nora.nerc.ac.uk/1084>
- Loudon, T.V. (2012). *A Scenario for Systems Geology*. BGS Research Report . www.bgs.ac.uk/systemsgeology
- NORA, 2010. The NERC Open Research Archive. <http://nora.nerc.ac.uk>
- Pepper, S., 2002. The TAO of topic maps. <http://www.ontopia.net/topicmaps/materials/tao.html>
- Procter, R., Williams, R., Stewart, J., Poschen, M., Snee, H., Voss, A., Asgari-Terghi, M., 2010. Adoption and use of Web 2.0 in scholarly communications. *Phil. Trans. R. Soc. A* **368**, 4039-4056.
- Sinclair, P., 2009. Linked Data on the BBC <http://www.slideshare.net/metade/linked-data-on-the-bbc>
- TopicMaps.Org Authoring Group, 2001. XML Topic Maps (XTM) 1.0: TopicMaps.Org. <http://www.topicmaps.org/xtm/index.html>
- Trippe, B., 2005. Component content management in practice: meeting the demands of the most complex content applications. The Gilbane Report: White Paper. http://gilbane.com/whitepapers/X-Hive/Xhive_Gilbane_Whitepaper_CCMS6-final.pdf
- Trippe, B., 2008. Using XML and databases: W3C standards in practice. The Gilbane Group, White Paper. <http://gilbane.com/whitepapers/EMC/EMC-XML-whitepaper-4.08.pdf>

Glossary

Codify: to create a representation or record of something in a form appropriate to the organised system of which it becomes a part.

Content: the persistent expression in physical form of intended meaning in information exchange, as opposed to its format or method of delivery.

Cyberinfrastructure: An integrated assemblage of computing, information and communication facilities, deploying the combined capacity of multiple sites to provide a framework to underpin research and discovery, typically with broad access and end-to-end coordination.

Configuration: The spatial arrangement, pattern, form and shape of objects, processes and their properties.

Data: A collection of observations, measurements or other information about a set of variables, generally in the form of numbers, words, or images. In the computer science sense, data is information suitable for use by a computer, distinct from the program (a set of instructions to perform a task).

Framework: A logical structure and guidelines giving a broad overview for classifying and organizing complex information, within which detail can be added as required.

Geographic Information System (GIS): An integrated system for the capture, storage, management, retrieval, analysis, manipulation and display of geographically referenced spatial data and its attributes.

GeoIndex: Search BGS datasets using GeoIndex; a map-based index of a wealth of information that we have collected or have obtained from other sources. <http://www.bgs.ac.uk/GeoIndex/home.html>

Geology: Geology is defined in the AGI Glossary as: "The study of the planet Earth – the materials of which it is made, the processes that act on these materials, the products formed, and the history of the planet and its life forms since its origin. Geology considers the physical forces that act on the Earth, the chemistry of its constituent materials, and the biology of its past inhabitants as revealed by fossils...The knowledge thus obtained is placed in the service of man – to aid in the discovery of minerals and fuels of value in the Earth's crust, to identify geologically stable sites for major structures, and to provide foreknowledge of some of the dangers associated with the mobile force of a dynamic Earth." (Bates and Jackson, 1980).

Geological Survey Organisation(GSO): "A GSO is a not-for-profit government organisation responsible for a range of tasks that generally include: geological surveying (mapping) of the nation, state, or province; conducting geological research to support economic development, public health, and environmental protection; distributing geoscience information; and advising government at various levels regarding water, mineral and energy resources, environmental issues, and earth hazards." (Berg et al., 2011).

Linked data: A style of publishing and interlinking structured data by looking up URI's on the Web. As opposed to seeing a fixed set of data sources, 'Linked Data' aims to create an unbound global data space. It links things in the world through their descriptions, by means of structured data on the Web.

Minimum revisable unit: A self-contained subset of information in a documentation system: a component of the system designed to be revised as necessary without endangering the integrity of the system as a whole.

Narrative: A message, story or discourse relating sequences of, for example, real or imagined events, descriptive features, or reasoning.

Ontology: A formal representation describing concepts, entities and relationships in a domain of knowledge, typically providing a more detailed and rigorous machine-readable specification than a thesaurus or taxonomy.

Scale-space: In GIS, scale is the ratio of the distance between two points on a map or image to the corresponding distance on the ground. In the study of scale-space, however, scale may refer to the level of detail detected by eye or instrument across a range of scale from the finest detail discriminated, to the entire image or field of view. Scale-space theory regards this range as a multi-resolution continuum, zooming in or out to reduce or increase the amount of detail using a filtering process. It studies the range of scales over which object types exist and geological processes operate.

System: A set of interacting parts that function as a whole. The systems approach involves study of linkages or interfaces between the component activities.

Systems geology: A view of geology re-based on the developing cyberinfrastructure and regarded as a system (a set of interacting parts that function as a whole) embedded in the wider knowledge system.

Uniform Resource Identifier (URI): A string of characters giving an address on the Web that identifies a physical or abstract resource (such as a thing, document, image, or service). A Uniform Resource Locator (URL) is a URI that provides the means of locating the resource.