

Developing and Managing Science Language for Enterprise Geoscience Data Management Systems

Peter H. Davenport

Geological Survey of Canada, 3303 33rd St NW, Calgary, Alberta

Peter.Davenport@NRCan.gc.ca

Consistent and clearly defined taxonomy is key to the successful management of geoscientific information in databases. Classifications traditionally used in Geoscience were not developed for use in data management, and many lack rigour in the meaning of terms, and consistency in their thematic organization. New classifications have been developed for data management to address these problems, but these commonly fail to gain acceptance among geoscientists because of their unfamiliarity and novel terminology. To bridge this gap, new ways of organizing traditional classifications for use in data management must be found.

Classifications used for geologic units (Lithostratigraphic, Lithodemic, etc.) are well defined by guidelines (e.g. the North American Stratigraphic Code), which make their use in data management straightforward. Once a Lexicon of such terms is constructed as a database, it becomes much easier to manage revisions and additions to the list of terms by applying the rules, and to engage efficiently experts to review proposed changes. Organizations may decide to relax certain of the rules of the stratigraphic code, permitting incompletely defined terms to be used informally in databases and publications, but still in a managed way. The Canadian Lexicon of Geological Names on-line http://www.cgkn.net/weblex/weblex_e.pl provides public access to the names of geological units and information about them. It also provides a channel for feedback that allows errors and gaps in the information to be submitted to the Lexicon database manager. Each lexicon term is cross-referenced against its originator, location (at the province/territory level) and age, allowing the lexicon to be searched by these criteria, as well as unit name.

The same general approach can be applied to other classifications that have few or no formal rules, for example terms for Earth Materials. There are some well developed, rule-based classifications for some groups of Earth Materials that are widely used, but at the other end of the spectrum, old terms that are from common language (e.g. shale, greenstone) are still commonly used by geologists. The meaning of some of these old terms lacks precision, especially when used out of context. To manage Earth Material terms for databases at the Geological Survey of Canada, a list of commonly used names (mostly in English) has been accumulated. These are checked against widely used Glossaries of Geoscience, from which definitions are derived. Terms not found to be in common use are excluded. For terms with divergent definitions, the one believed to reflect common usage is selected. From these definitions, each Earth Material term is cross-referenced against a set of simple criteria (including genesis, composition and texture), enabling searching.

Both the Lexicon and Earth Material databases are used as pick lists for geological mapping and sampling at the Geological Survey of Canada. Requests for the addition of terms not in the databases are submitted to the database manager, and are added after a simple review process. The database is thereby updated continuously. It is anticipated that just as geological units are revised, and some fall from use, the same process can be applied to Earth Material names (and other types of science language), with some of the older, loosely defined terms falling from use as better defined terms replace them. Furthermore, a selection of terms from the complete list may be made for use in a particular project or purpose, providing another way to exclude obsolete terms.