Controlled Vocabulary Design Draft Plan

Learn about the System Infrastructure of IEDA

SESAR

SESARDB

EarthChem Portal

Internal (Controlled): PetDB SedDB VentDB DLDS

External (Uncontrolled): MGDS Geochron MAGA GVP
NavDat USGS   MetPetDB
GANSEKI GEOROC ODM2   ...

EarthChem Library

ECLDB

GeoMapApp

Question: Will EarthChem DB replace PetDB, SedDB, VentDB and DLDS?

Clarify the purpose of building CV

What will the CV be used for?

The structure of CV depend on the purpose.

Indexing/Tagging support

example: one sample data may be indexed by sample type, rock
classification, etc.

Retrieval support (including both search and discovery)

Organization and Navigation support

Define the scope of indexed/retrieved content

Limit the content to DB, exclude other media like image, paper, etc..

Which db will be indexed by CV?

Internal DB like PetDB

Which db will be retrieved using CV?

Internal DB and external DB

What types of subjects can be included in the CV?

Check the CV of distinct DB (including internal and external).

Check the metafield of the distinct search interfaces across different systems.

Create CV

- Determine which format will be used for CV.

  SKOS? OWL? **(Prefer to SKOS)**

- Determine which structure will be used for CV.

  Flat? Hierarchical? Facets? **(Maybe hybrid)**

  Hierarchies: **have one dimension of classification, divide and redivide things into groups where each new group is a sub-groups of its parent group.**

  Facets: handle three or more dimensions of classification.

- Identify the top categories or facets based on analysis and scope of concepts.

  geographic features | physiographic features | tectonic settings

- Gather term concepts from content audit, interviews, and other sources.

- Come up with example second-level terms for each category (possibly even a few third-level terms).

- First review and revisions (internal, meeting).

- Making full use of all the concepts gathered, fully build the CV out to lower levels.

- Second review and revisions (internal, meeting).

- Import them into CV management software.

- Create non-preferred term, scope notes or other term details (if necessary).

- Third review and revisions (internal, online)

- Test CV with sample indexing and retrieval, and revise it (extra coding work).

- Put it to public, accept review from outside of IEDA, and revise as needed.

Standards

ANSI/NISO Z39.19-2005 (R2010)

Software for CV creation and management (open source stack)

SISSVoc (Java servelet)
iQvoc (Ruby & Rail)
TemaTres (LAMP)
what we need to consider when we choose a CV software?
Stable URIs, SPARQL endpoints, web service, capability of management, capability of publishment, database server supporting

Prepare CV Server

applying domain name

Setting up CV server

Other issues

Develop new interface for CV management software to keep consistency with IEDA web Style (if needed, extra coding work)

Port CV to distinct DB which need to be indexed (extra coding work, hard part)

CV maintenance (discuss in future)

CrossRef or Validation with external CVs

GeoRef Thesaurus Lists http://www.americangeosciences.org/georef/georef-thesaurus-lists

List A   GeoRef Categories/Subjects covered

List B   Age dating methods

List C   Commodities

List D   Chemical hierarchies, methods, and data terms

List E   Geologic age (stratigraphic) terms

List F   Fossils

List G   Meteorites

List H   Igneous rocks

List I   Sedimentary rocks

List J   Metamorphic rocks

List K   Sedimentary structures

List L   Minerals

List M   Soils

List N   Sediments

List O   Geographic terms

List P   Major terms

List R   Rock units

Geologic age

the Geological Time Table, compiled by B. U. Haq, Sixth revised, enlarged and updated edition, Amsterdam, Elsevier, 2007.

the International Commission on Stratigraphy's International Chronostratigraphic Chart

http://www.stratigraphy.org/index.php/ics-chart-timescale

updated each year

Minerals

IMA List of Minerals at IMA-Mineralogy   http://www.ima-mineralogy.org/

updated every month

Rocks classification

Igneous rocks; a classification and glossary of terms, edited by R. W. Le Maitre, second edition, New York, Cambridge University Press, 2002

variables (parameters, items)

Chemical Abstract Service (CAS) Registry: http://www.commonchemistry.org

Chemical Entities of Biological Interest (CheBI): http://www.ebi.ac.uk/chebi/

Methods

National Environmental Methods Index (NEMI) https://www.nemi.gov/home/