A generic approach to manage metadata standards

Barde Julien[1], Edgington Duane[1], Desconnets Jean-Christophe[2]

[1]Monterey Bay Aquarium Research Institute (MBARI)

[2]IRD, US ESPACE, Maison de la télédétection

Wednesday, September 26

Outline

Table of contents

Improvement of data processing and decision-making

Locating exhaustively the relevant Informational Resources (IR) is a challenge:

- IR are heterogeneous (*language*, *metadata*, *syntax/formats*, *semantic*, *access constraints*...),
- IR are distributed into heterogeneous Information Systems (IS) whose interoperability first involves matching issues.

Institut de recherche pour le développement

Key issue: {metadata element, value} pairs

Once aware of existing IS, the priority to locate the related IR consists of managing the matching between the heterogeneous metadata elements (syntactic) as well as between their heterogeneous values (semantic).

Metadata management still faces the lack of referentials

Heterogeneous metadata elements sets interfere with IS syntactic interoperability:

- similar core metadata elements (where? what? when? who?...) but different *terms* generate matching issues,
- new international metadata standards make the previous ones obsolete (because of redundant scopes) and bring archiving issues,
- recent use of XML Schemas to standardize their implementation (standards redundancy decreases as they are used as *types libraries*).

Whatever the standard, heterogeneous values of *core metadata elements* interfers with IS semantic interoperability: the use of *common semantic and geographic/spatial referentials* is a key issue.

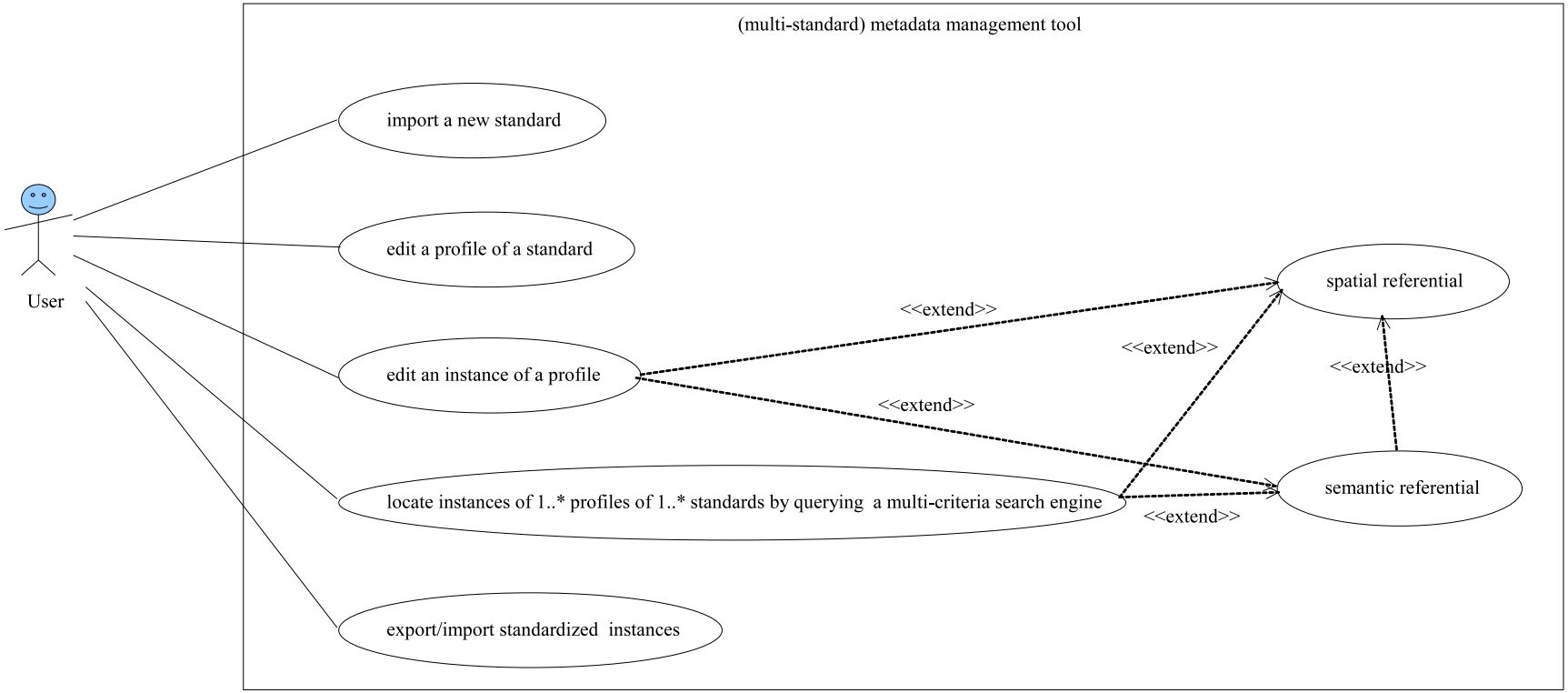
Users tasks and needs

Users tasks are similar from one standard to another:

- profiling the metadata standards for their specific uses,
- editing *standards instances* (by relating values to profiles metadata elements),
- locating IR described with different standards by using a single *multi-criteria search engine*,
- import/export *standards instances* (usually XML).

Users need assistance to perform these tasks easily : a single centralized access, a tool with homogeneous and friendly GUIs assisting any complex valuation process (*Web mapping tool*, *controlled vocabularies*, *calendar*...).

Users tasks and needs



Software engineers tasks and needs

Software (engineers) tasks consist of:

- satisfying user needs by complying with standards,
- matching *terms* related to similar *core metadata elements* to answer basic multi-standard queries,
- integrating semantic and spatial referentials to control related values,

Software (engineers) need to minimize their developements efforts (by answering the same user needs in the same way): *generic approach*. Whatever standard implemented: a single scripts set and the same components (WMS. . .).

The good, the bad and the ugly

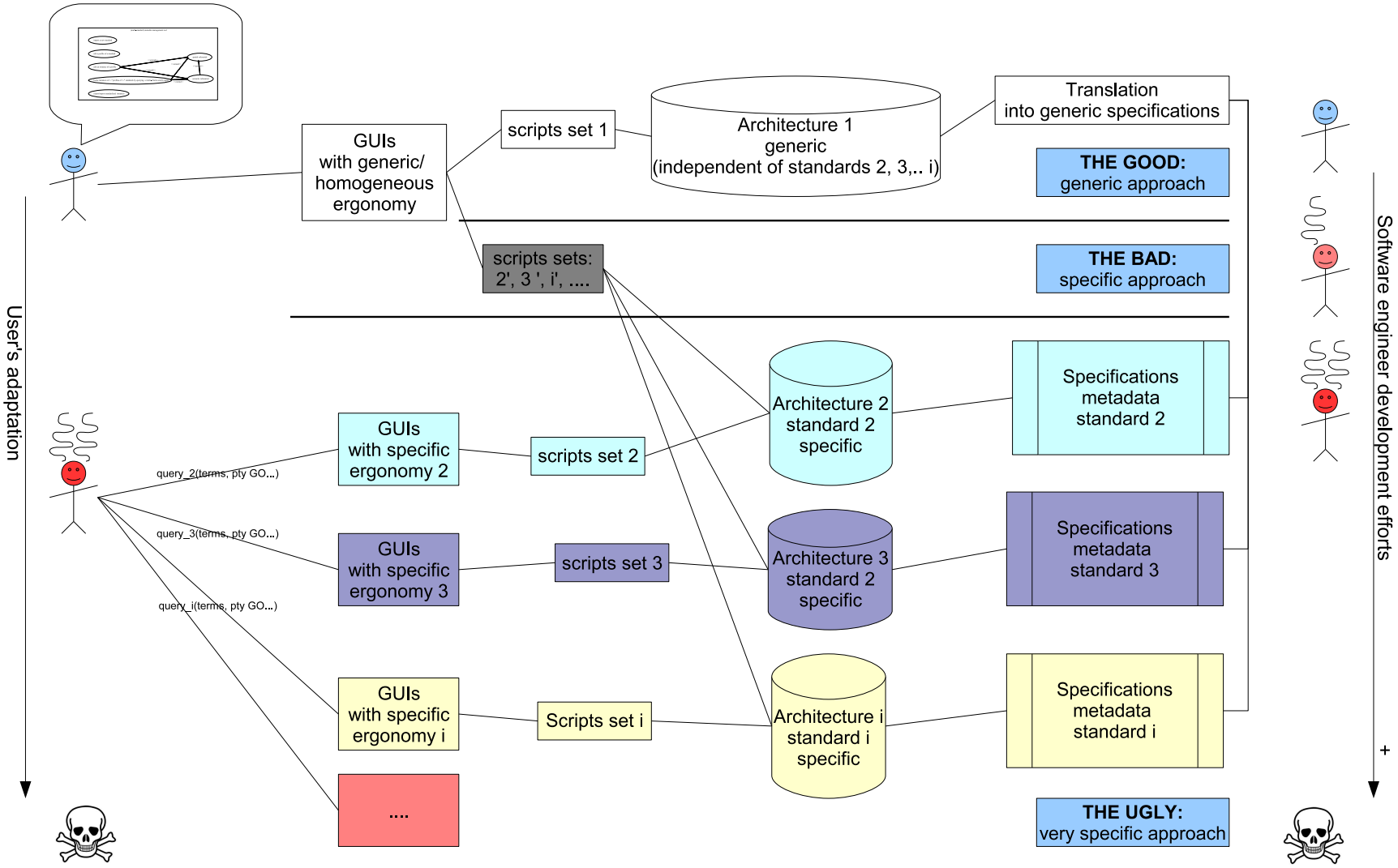
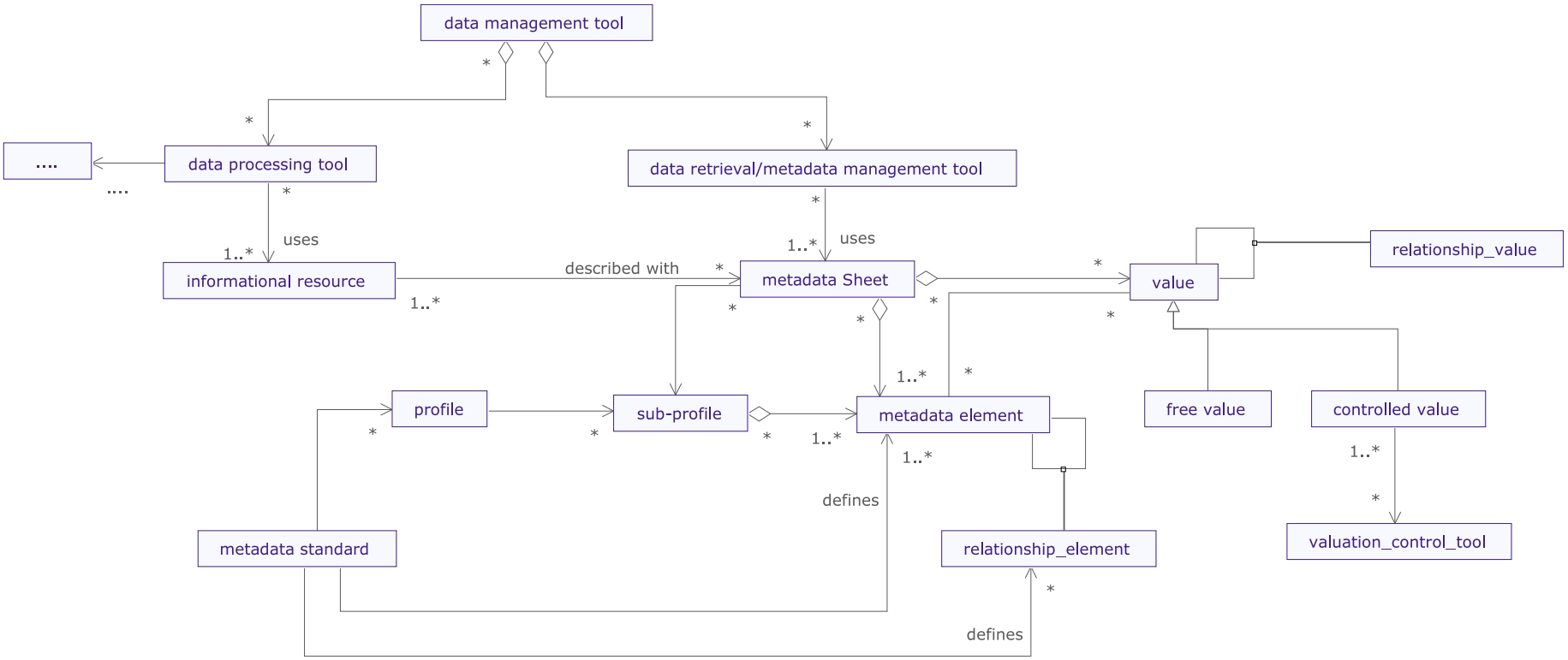


Table of contents

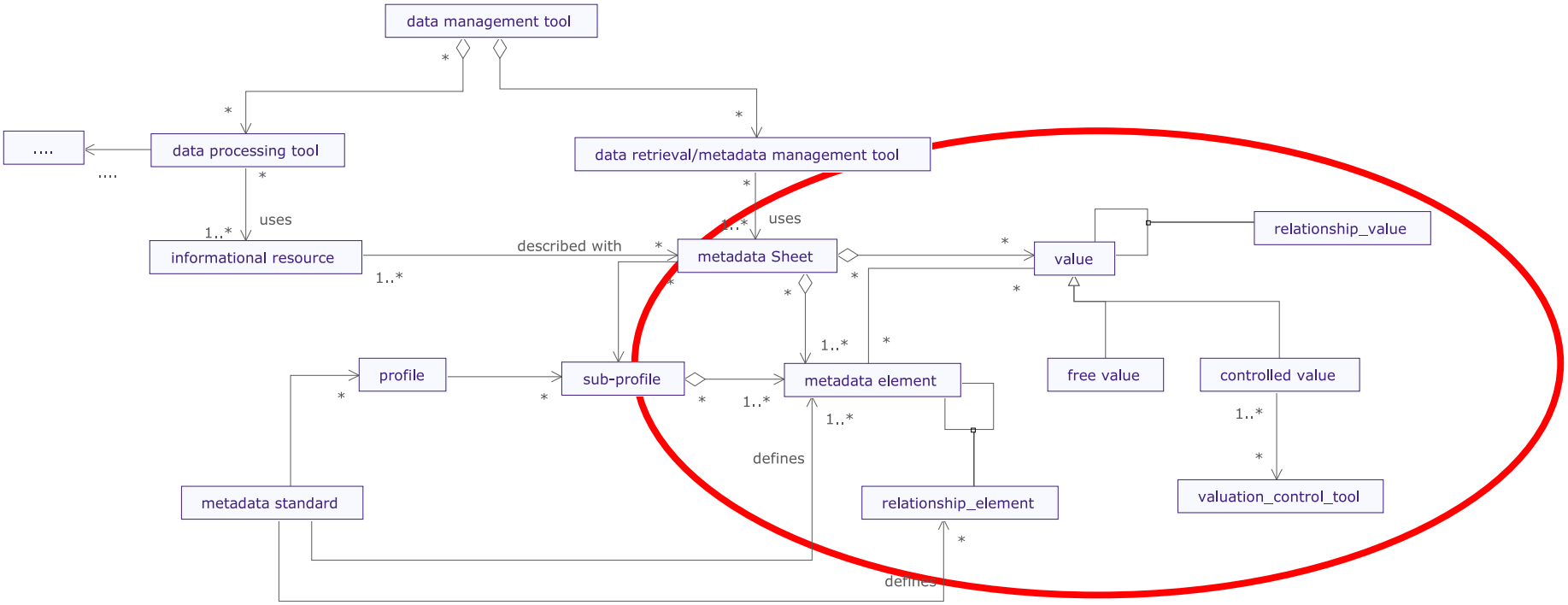
Conceptual model for multistandard metadata management



Goal

A generic pattern (conceptual model) to describe any MD standard and then set up a multistandard metadata management tool assisting valuation. {metadata element, value}

Conceptual model for multistandard metadata management



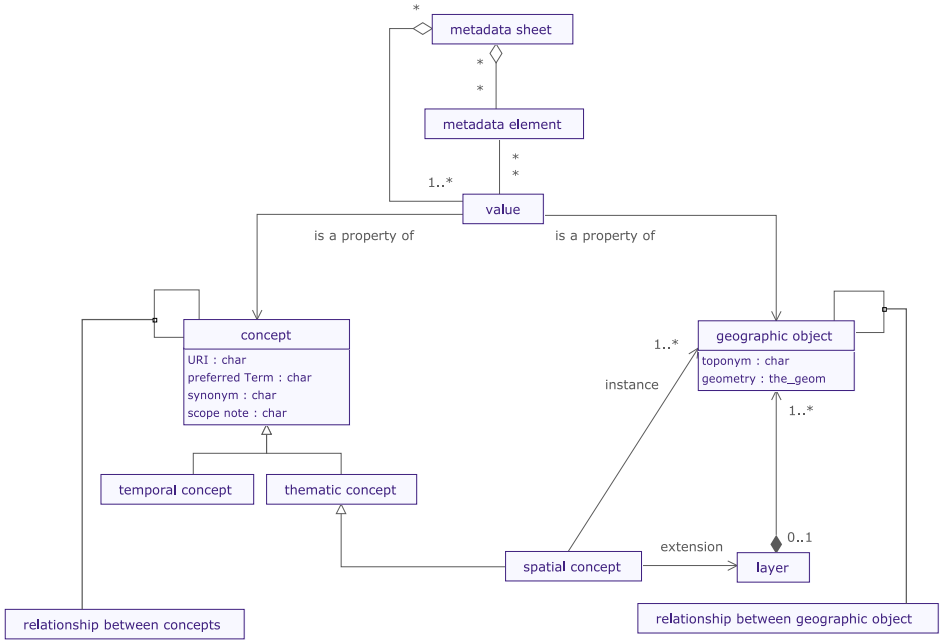
Goal

A generic pattern (conceptual model) to describe any MD standard and then set up a multistandard metadata management tool assisting valuation. {metadata element, value}

Relationship between spatial and thematic concepts

Use of terms in core metadata elements is often ambiguous:

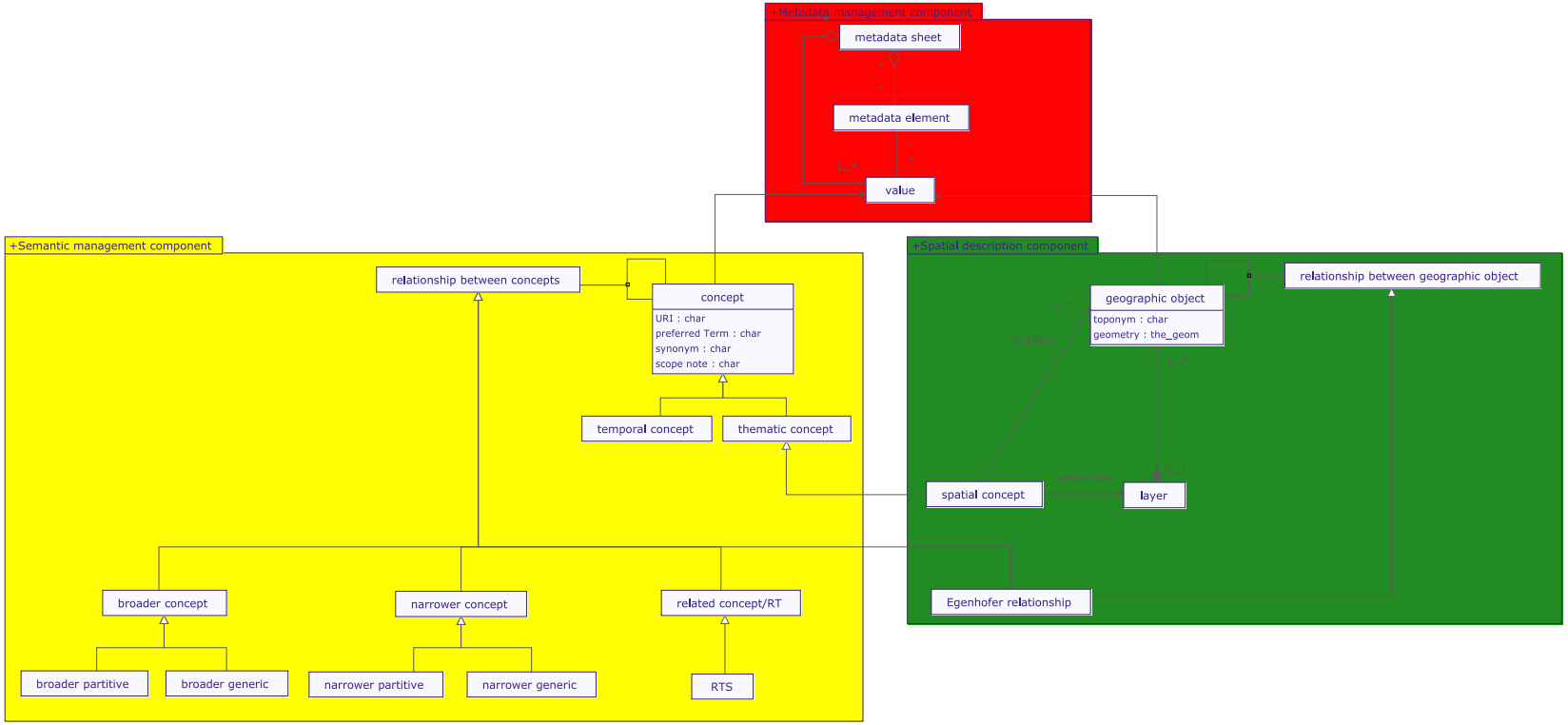
- Example: "*swordfish, sea temperature, Madagascar, spring*"
- Semantic relationships management: synonyms ("*Xiphias gladius...*"),  ...,
- The case of toponyms: keywords or spatial description?  ...



Relationship between thematic and spatial descriptions

We make explicit the relationship between thematic and spatial concepts: "*a spatial concept as a kind of thematic concept whose instances are geographic objects*".

A generic model to link between MD and ontology

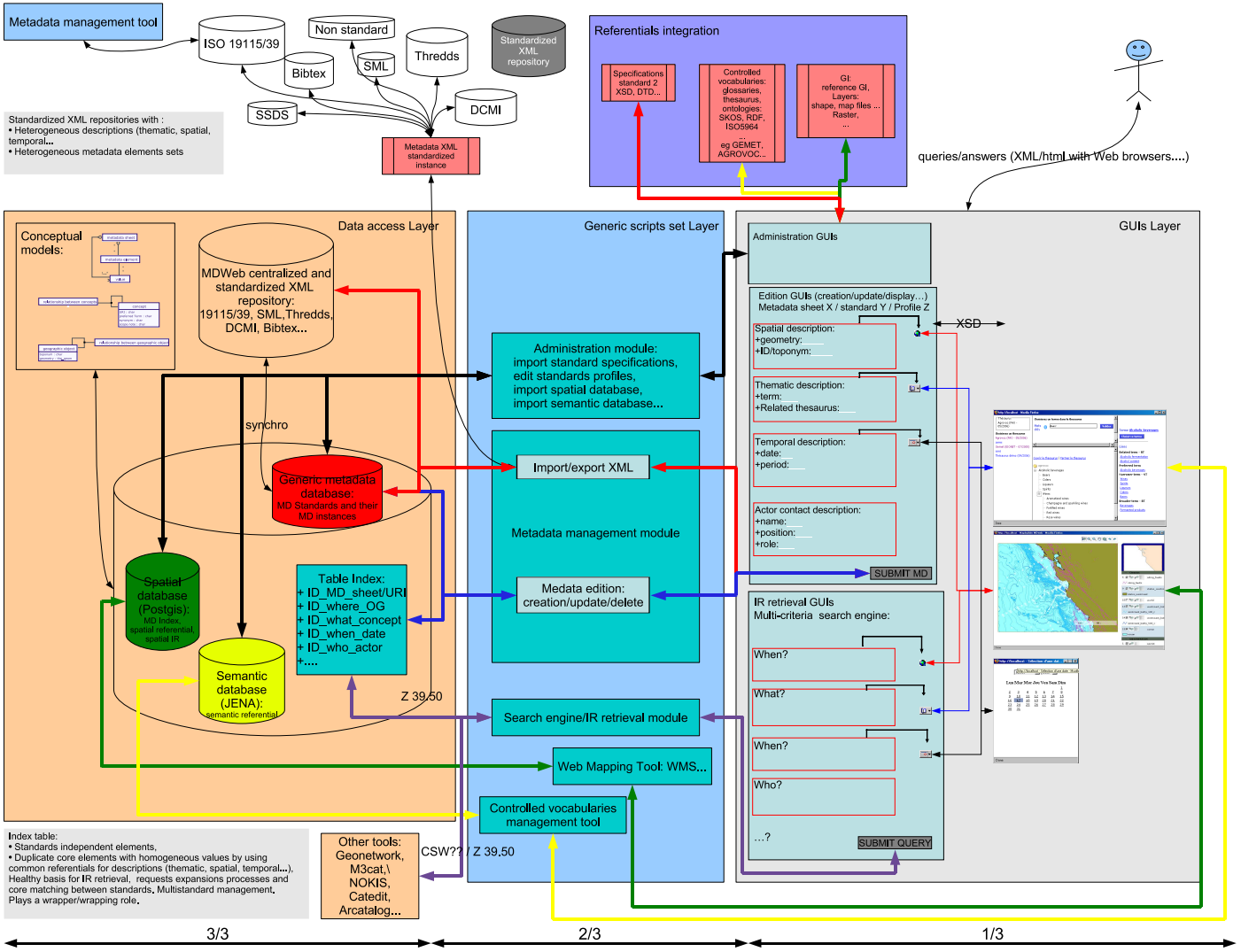


Compliance with standards

Model compliant with standardized implementations of metadata standards (*XML Schemas*, *DTD*...), (Web) Semantic standards *SKOS* (*ISO 2788/5964*)/*RDF*/*OWL*, main GI standard formats.

Table of contents

Architecture of the physical data infrastructure



Related open source software

Current implementations based on MDWeb open source software:

MDweb Open source CeCILL 1.5
Tool for cataloging and locating environmental information

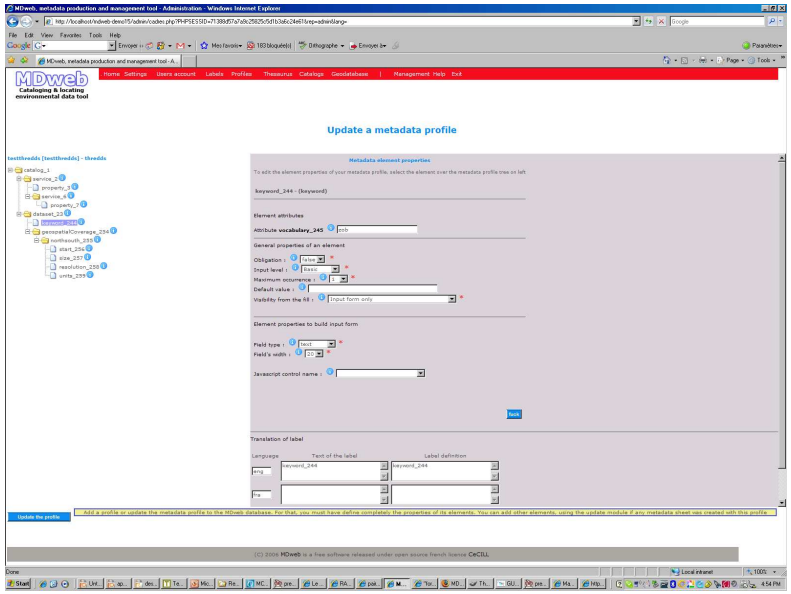
http://www.mdweb-project.org/

- a *multistandard* and *multilingual* metadata cataloging tool implementing a *generic approach* (like *M3Cat*, *MetaCat*...),
- a three-tier (client-server) architecture with:
  1. GUIs (in Web browsers) with components to assist metadata edition or searching:
     - *spatial description* with Web Mapping tools: Mapserver / Mapbuilder,
     - *thematic description* with Controlled vocabularies tool: home made.
  2. Applications scripts: PHP/Javascript/XML with Apache Http server ,
  3. Data storage: *RDBMS*: Postgres with Postgis . Import of SKOS files into Postgres by using JENA java API. *XML repositories*.

Same GUIs regardless of standards

MDWeb used as a basis to allow:

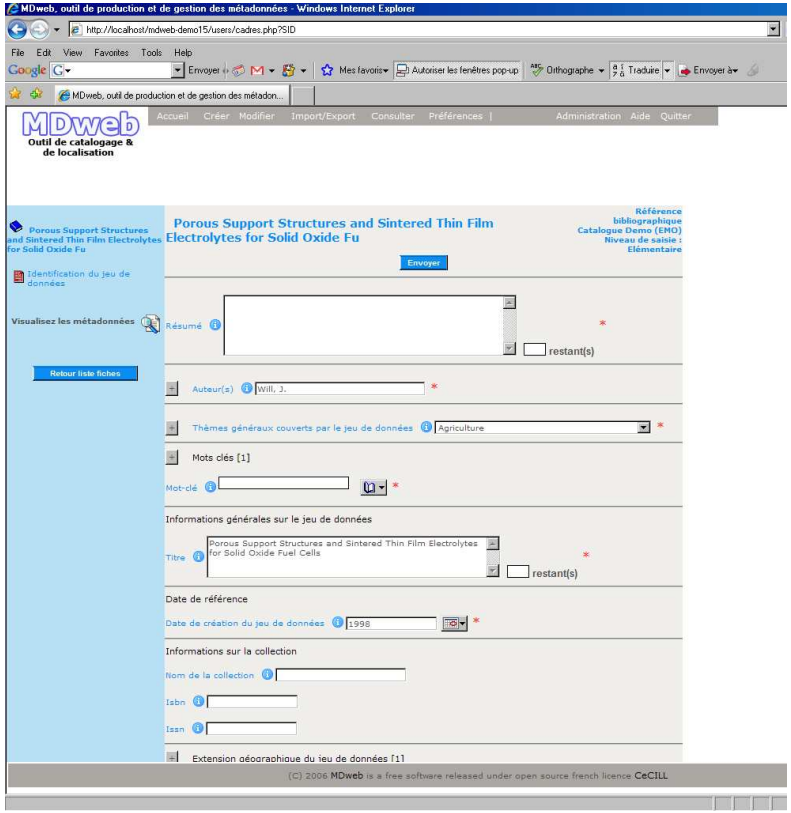
- Import of any new metadata standard by translating formal specifications into the PDM (automated with XML schemas),
- Set up of profiles of imported standards,



Same GUIs regardless of standards

MDWeb used as a basis to allow:

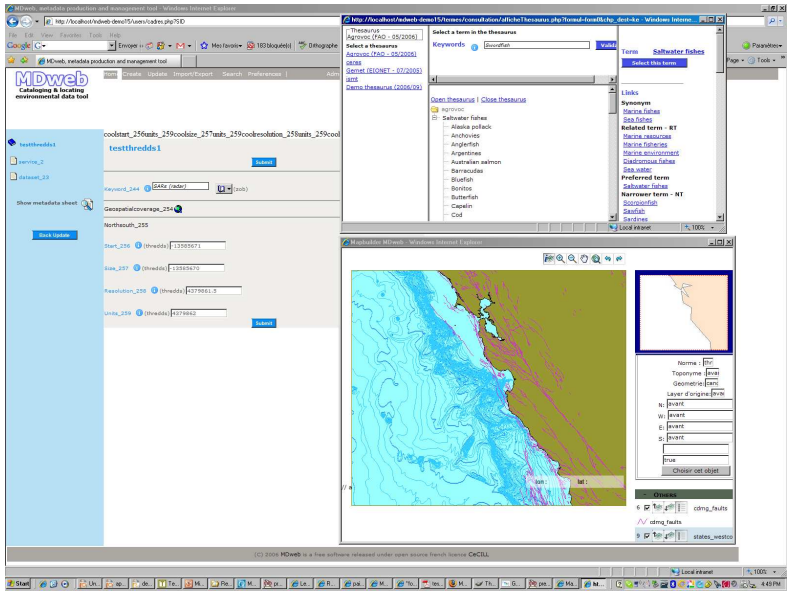
- Import of any new metadata standard by translating formal specifications into the PDM (automated with XML schemas),
- Set up of profiles of imported standards,
- Metadata sheet edition with additional GUIs to assist RI thematic and spatial descriptions (automatizations, control. . .),



Same GUIs regardless of standards

MDWeb used as a basis to allow:

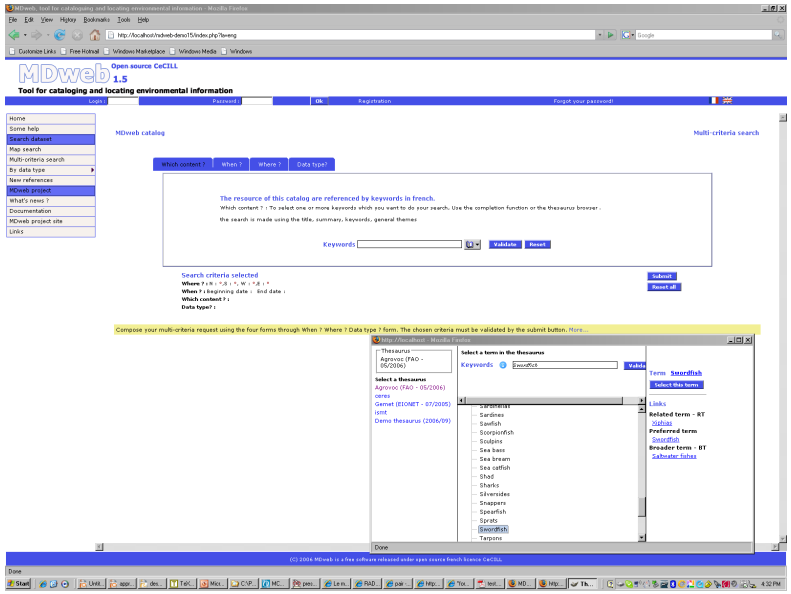
- Import of any new metadata standard by translating formal specifications into the PDM (automated with XML schemas),
- Set up of profiles of imported standards,
- Metadata sheet edition with additional GUIs to assist RI thematic and spatial descriptions (automatizations, control. . .),



Same GUIs regardless of standards

MDWeb used as a basis to allow:

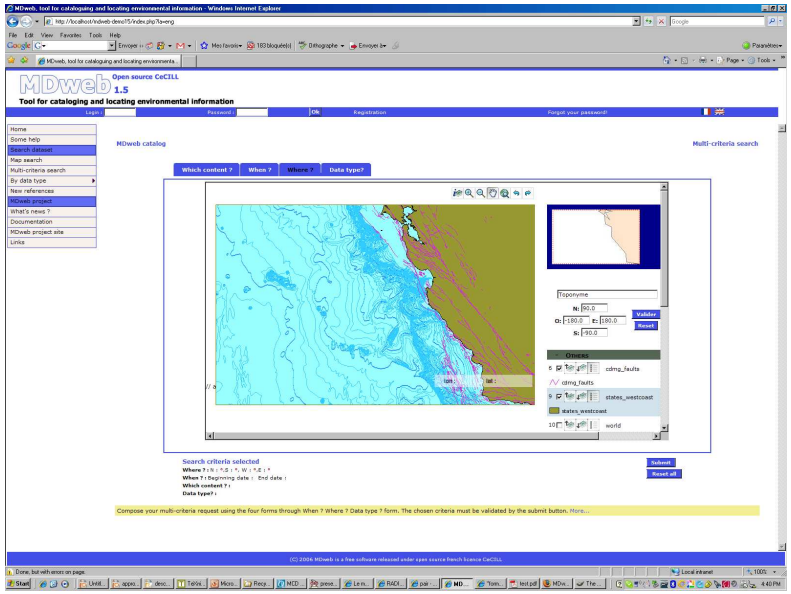
- Import of any new metadata standard by translating formal specifications into the PDM (automated with XML schemas),
- Set up of profiles of imported standards,
- Metadata sheet edition with additional GUIs to assist RI thematic and spatial descriptions (automatizations, control. . .),
- Generic/multi-standard search engine,



Same GUIs regardless of standards

MDWeb used as a basis to allow:

- Import of any new metadata standard by translating formal specifications into the PDM (automated with XML schemas),
- Set up of profiles of imported standards,
- Metadata sheet edition with additional GUIs to assist RI thematic and spatial descriptions (automatizations, control. . .),
- Generic/multi-standard search engine,



Same GUIs regardless of standards

MDWeb used as a basis to allow:

- Import of any new metadata standard by translating formal specifications into the PDM (automated with XML schemas),
- Set up of profiles of imported standards,
- Metadata sheet edition with additional GUIs to assist RI thematic and spatial descriptions (automatizations, control. . .),
- Generic/multi-standard search engine,
- Import/export of standardized XML metadata sheet.

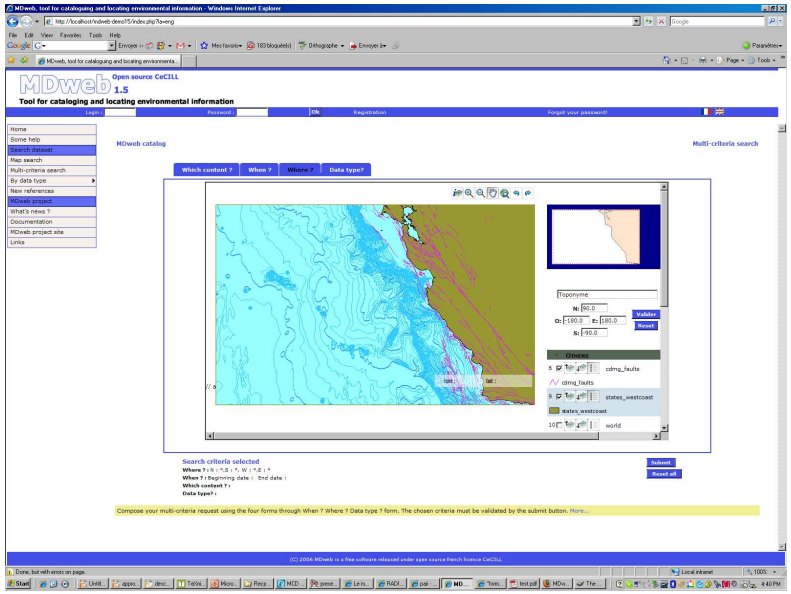


Table of contents

Conclusion

A single generic architecture based on *standards and opensource software* can manage:

- heterogeneous metadata standards (import, profiling, edition...),
- heterogeneous values: in particular *controlled terms* and *spatial descriptions* to describe *core metadata elements*,
- a common indexation table duplicating *core metadata elements* with homogeneous *terms* and *values* used by the *search engine* (no wrapper needed),
- spatial IR described by metadata can be processed after being retrieved: either locally or remotely by using interoperable protocols or/and rich clients (*WMS*, *Qgis*, *Udig*...).

Data retrieval is thus improved, in particular queries expansion by using standardized semantic or spatial relationships (requests portability).
Both *user's and software engineer's tasks and needs* are taken into account by a *single scripts set*.

Outlook

Outlook for the *metadata management* :

- complete the implementation of all use cases:
  - generic metadata standards import and edition of profiles and instances,
  - generic / multi-standard metadata search engine,
  - generic standardized XML import and export.
- improve both current spatial queries and requests expansion process by using richer semantic and spatial relationships.

Outlook for *data processing*: use the current architecture to homogenize formats and semantic (just like metadata. . .) and then display (plot. . .), subset / aggregate. . . with ability to use interoperable rich clients instead of light clients.