



VMF (Vocabulary Mapping Framework) Project Final Report, Godfrey Rust, December 12 2009

Acknowledgements

The Project was funded by JISC. The active co-operation and support of a number of major metadata-related initiatives – especially ONIX, RDA, Dublin Core, DDEX, IDF, FRBR, all of whom are represented in the ongoing Advisory Board – has been essential for the progress made, and the individual members of the Advisory Board are thanked for their time and input. The British Library has supported the initiative in its original phase (the RDA/ONIX Framework) and in this, including hosting the final seminar: again their contribution has been invaluable.

Executive Summary

The goal of the Project was simple: to establish an ontology (or structured dictionary of terms) in which different terms from controlled vocabularies in a wide range of content-related metadata standards could be mapped together for the primary purpose of *automatically* generating mappings between them to support reuse of metadata.

At present such mappings are produced in a variety of ad hoc, one-to-one piecemeal initiatives. The rapid growth and rate of change among metadata schemes, both standard and proprietary, has made it clear that it is responsible organizations that are to be able to effectively and accurately reuse other people's data, as is essential in the networked, collaborative digital content environment, that some tools for orderly and reliable transformation are needed. Because such mappings require authoritative understanding of the precise meanings of terms in the schemes being mapped, such tools could not easily be produced and maintained unilaterally.

The tool produced would be freely available – and is now available via the project website at <http://cdlr.strath.ac.uk/VMF/>. This was seen as an achievable and practically useful short/medium-term target which could also provide a focal point for the development of other interoperability tools among the metadata schemes involved. The longer term outcome is expected to be the ongoing maintenance of the ontology (the “VMF matrix”) as an authoritative tool for vocabulary mapping and development with potential to bring significant benefits to the bibliographic and content publishing communities. Other processes can be built on it, and there is no inherent limit to its scope in terms of content.

The matrix was built from “prior art”, drawing on the work of earlier projects, especially the <indec> framework and the RDA/ONIX Framework. The technical requirements were kept as simple as possible to concentrate on the semantic issues. The outputs were produced in the Semantic Web standards (RDF/OWL) in a common and accessible format.

The process was simple in principle: a rich “hub” ontology would provide all the necessary points to which vocabulary terms could be mapped, and in doing so provide a matrix of internal semantic links to relate them together, so that by writing queries in a standard language such as SPARQL the “best fit” mappings could be produced between any chosen pair of vocabularies.

The technical issues of mapping diverse terminology are not trivial, and this first stage project was a “proof of concept” which would create a tool that was real but limited and not fully tested or “debugged”: its aim would be to establish a methodology which could be seen to work. Stage 2 would turn this from an “alpha” to a beta or even production release.

A small team (three) carried out the work, with excellent support from an Advisory Board. The technical work was carried out by two Rightscom consultants, who drew on a great deal of prior work

Project Acronym:
Version:
Contact:
Date:

on related projects (including work on several of the standards being mapped), so the amount of genuine R&D was limited. The hub ontology is based on an event-based data model used successfully in several existing standard and proprietary schemes and systems. This model is the technical key to the matrix.

The results appear good. No major unexpected issues were encountered. Approaching one thousand terms from a diverse range of more than 50 vocabularies have been successfully mapped, and some exemplary automated mappings produced (stage 1 was mainly about input – stage 2 will deal more thoroughly with “output”). A number of new approaches were tried, apparently successfully, for dealing with specific issues. A number of further steps were identified for implementation in the next stage. The main concern – that the matrix would grow to be unmanageably complex – has so far not proved to be a concern. Working with widely available tools (Excel and Protégé) it is very manageable, and the outputs are processible by anyone using an OWL or similar reasoner.

The final seminar confirmed a significant level of interest in the VMF, and the will among the participating schemes to turn it into a ongoing tool.

Background

The background to the project is given in the project proposal (also available via <http://www.jisc.ac.uk/whatwedo/projects/vocab-framework.aspx>)

Aims and Objectives

The aim of the project was: to establish an ontology (or structured dictionary of terms) in which different terms from controlled vocabularies in a wide range of content-related metadata standards could be mapped together. for the primary purpose of *automatically* generating mappings between them to support reuse of metadata. This did not change in the course of the project.

Methodology and implementation

The method of building the matrix is described in summary and in detail in the delivered document “The Vocabulary Mapping Framework: an introduction” - [http://cdlr.strath.ac.uk/VMF/documents/VocabularyMappingFrameworkIntroductionV1.0\(091212\).pdf](http://cdlr.strath.ac.uk/VMF/documents/VocabularyMappingFrameworkIntroductionV1.0(091212).pdf)

Outputs and Results

The outputs and results are described in the delivered document “The Vocabulary Mapping Framework: an introduction”. The alpha version of the VMF matrix is published in an RDF file on the VMF website.

Outcomes, conclusion, implications

The project met its stated objectives, and with a further stage of development should be usable by the JISC and related communities as a practical tool to support the transformation of metadata between different schemes and databases.

The potential for other uses of the VMF framework, for example:

- as an mediator for searching/querying across multiple data formats
- for taxonomy/subject index mapping
- for full message transformation

is considerable, as the underlying semantic issues are the same and are not being seriously addressed on any scale anywhere else.

The establishment of an ongoing governance mechanism, grown from the current Advisory Board, can also provide a unique point of interaction between the major metadata initiatives in these sectors to support other developments like those referred to. It is the first time such a wide range of content metadata schemes have supported a collaborative interoperability project.

The Advisory Board is exploring possibilities for funding for the next stage of development from several sources. This includes further funding from JISC. While we are aiming for VMF to be a

Project Acronym:
Version:
Contact:
Date:

permanent service, we are not expecting JISC to support any more than one further stage of development at the most.

References

Details of sources and other references are given on the VMF website - <http://cdlr.strath.ac.uk/VMF/>