Software for Managing and Sharing Metadata Translation Information

Mark Hedley
Met Office
United Kingdom

September 24, 2014
Three Steps to Interoperability

- Publishing standards:
  - Publishing standards: accessibility to specific concepts: for humans; for machines.
  - Relating standardised terms: http://metarelate.net
  - Translating data sets: http://scitools.org.uk/iris
Three Steps to Interoperability

- **Publishing standards:**
  - accessibility - to specific concepts:
    - for humans;
    - for machines.
Three Steps to Interoperability

- **Publishing standards:**
  - accessibility - to specific concepts:
    - for humans;
    - for machines.

- **Relating standardised terms:**
  - [http://metarelate.net](http://metarelate.net)
Three Steps to Interoperability

- Publishing standards:
  - accessibility - to specific concepts:
    - for humans;
    - for machines.
- Relating standardised terms:
  - http://metarelate.net
- Translating data sets.
  - http://scitools.org.uk/iris
What is Linked Data?

- Linked Data describes a method of publishing structured data so that it can be interlinked and become more useful.
- Three key steps to linked data:
  - use open standards to publish information,
  - use URIs to identify resources,
  - link to other people’s resources.
What is Linked Data?

- Linked Data describes a method of publishing structured data so that it can be interlinked and become more useful.

- Three key steps to linked data:
  - use open standards to publish information,
  - use URIs to identify resources,
  - link to other people’s resources.

- There are semantic web technologies and standards to help us; e.g.:
  - RDF, OWL, SKOS, SPARQL;
  - http://www.w3.org/standards/semanticweb/data
What is Linked Data?

- Linked Data describes a method of publishing structured data so that it can be interlinked and become more useful.

- Three key steps to linked data:
  - use open standards to publish information,
  - use URIs to identify resources,
  - link to other people’s resources.

- There are semantic web technologies and standards to help us; e.g.:
  - RDF, OWL, SKOS, SPARQL;
  - http://www.w3.org/standards/semanticweb/data

- Technologies can facilitate,
  - if we know what we want to do.
Serving CF Semantics

- Numerous communities, including CF, have been publishing semantics as linked data for a number of years.

- http://vocab.nerc.ac.uk/collection/P07/current/CFSN0015/
- or a locally downloaded RDF/XML file
- Courtesy of the British Oceanographic Data Centre and the British Atmospheric Data Centre.
Serving WMO Semantics

- IPET-MDRD are working to publish the manuals on codes as linked data:
  - http://codes.wmo.int
  - http://codes.wmo.int/grib2/codeflag/4.2/0-3-0
- or a locally downloaded one
- Courtesy of the WMO and the Met Office.
Publishing Standards

Introduction

Publishing Standards

Relating Terms

Translating Data Sets

In Summary
Publishing Mappings
Publishing Mappings

Introduction

Publishing Standards

Relating Terms

Translating Data Sets

In Summary
Responsibility, Trust

- Individual standards communities have responsibility for their domains and the semantics published within those domains.
  - Who owns these mappings?
  - Are they authoritative?
  - Can we trust them?
Responsibility, Trust

- Individual standards communities have responsibility for their domains and the semantics published within those domains.
  - Who owns these mappings?
  - Are they authoritative?
  - Can we trust them?
- To build trust, we need to develop:
  - a community to maintain the information;
  - delegated authority for this community to act with multiple standards bodies;
  - an information curation strategy and work flow;
  - a validation process for trusted information.
Metarelate

http://metarelate.net

© Crown copyright. Met Office
Metarelate is a free and open source collaboration project targeting metadata interoperability.

- GRIB and CF are two of the formats targeted for initial development.

It provides a potential place for standards communities and their users to collaborate on sharing and managing translation information.
Metarelate

- Metarelate is a free and open source collaboration project targeting metadata interoperability.
  - GRIB and CF are two of the formats targeted for initial development.

- It provides a potential place for standards communities and their users to collaborate on sharing and managing translation information.

- This may be able to be a focal point for a set of collaborative activities.
Software and Information

- Under the Metarelate domain there are two projects:
  - metarelate: a python API for managing metadata translation information;
    - built on semantic web technologies: Jena, Fuseki, SPARQL.
  - MetOcean: a database of mapping entries in the meteorology and oceanography domain.
Software and Information

- Under the Metarelate domain there are two projects:
  - metarelate: a python API for managing metadata translation information;
    - built on semantic web technologies: Jena, Fuseki, SPARQL.
  - MetOcean: a database of mapping entries in the meteorology and oceanography domain.

- Metarelate’s MetOcean translation database is currently a software development project only, with no live hosting of information.
Mappings

Introduction
Publishing Standards
Relating Terms
Translating Data Sets
In Summary

Software for Managing and Sharing Metadata Translation Information

Mark Hedley
Met Office
United Kingdom

Metarelate

Source Concept
Target Concept

© Crown copyright. Met Office
Complications

Metarelate

Source Concept

Target Concept

© Crown copyright. Met Office
Complicated Mappings

Introduction
Publishing Standards
Relating Terms
Translating Data Sets
In Summary

[Diagram showing mappings between concepts and metadata]

Mark Hedley
Met Office
United Kingdom

© Crown copyright. Met Office
Translating Data Sets

- We would like to make it easy for domain experts to contribute to the contents of the translation database.
- People need to see that they can:
  - use the information for something useful;
  - trust the information available.
We would like to make it easy for domain experts to contribute to the contents of the translation database.

People need to see that they can:

- use the information for something useful;
- trust the information available.

This is key to encouraging the engagement of specialist expertise required.
Using Metarelate Mappings

- Metarelate provides no data set translation capabilities.
Using Metarelate Mappings

- Metarelate provides no data set translation capabilities.
- Metarelate has attempted to be implementation neutral.
Using Metarelate Mappings

- Metarelate provides no data set translation capabilities.
- Metarelate has attempted to be implementation neutral.
- We can configure our software to use the Metarelate mappings.
Using Metarelate Mappings

- Metarelate provides no data set translation capabilities.
- Metarelate has attempted to be implementation neutral.
- We can configure our software to use the Metarelate mappings.
- This provides a potential path towards consistency of interpretation between different applications.
SciTools - Iris

Iris is a free and open source Python library which provides metadata translation capabilities between CF-netCDF and GRIB2 files (amongst others).

- Iris uses Metarelate as a trusted source of metadata translation information.
- Translations are sourced as a code change and provided in a packaged release.

© Crown copyright. Met Office
Conclusions

- We find that using the linked data approach works.
Conclusions

- We find that using the linked data approach works.
- Managing the relationships between different standards requires inter-community effort and shared responsibility.
Conclusions

- We find that using the linked data approach works.
- Managing the relationships between different standards requires inter-community effort and shared responsibility.
- We need domain experts to participate, contribute, use and benefit.
Next Steps?

- Assemble, collate and validate mappings between CF and GRIB2 terminologies.
Next Steps?

- Assemble, collate and validate mappings between CF and GRIB2 terminologies.
- Encourage the formation of a community of stakeholders to curate the information.
Next Steps?

- Assemble, collate and validate mappings between CF and GRIB2 terminologies.
- Encourage the formation of a community of stakeholders to curate the information.
- Develop a community translation curation process to build trust in the provided translations.
Next Steps?

- Assemble, collate and validate mappings between CF and GRIB2 terminologies.
- Encourage the formation of a community of stakeholders to curate the information.
- Develop a community translation curation process to build trust in the provided translations.
- Develop Metarelate’s MetOcean as a web service.
Questions?

Comments?

Thoughts?