Met Office

Describing
Statistical
Process
Results

Mark Hedley -
Analysis,
Visualisation
and Data
Team - UK
Met Office

Objectives

Opportunities

Challenges

Examples

Conclusions

Describing Statistical Process Results

Mark Hedley - Analysis, Visualisation and Data Team - UK Met Office

January 20, 2016

Outline

Describing
Statistical
Process
Results

Mark Hedley -
Analysis,
Visualisation
and Data
Team - UK
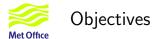Met Office

Objectives

Opportunities

Challenges

Examples

Conclusions

1 Objectives

2 Opportunities

3 Challenges

4 Examples

5 Conclusions

Objectives

Describing
Statistical
Process
Results

Mark Hedley -
Analysis,
Visualisation
and Data
Team - UK
Met Office

Objectives

Opportunities

Challenges

Examples

Conclusions

- Explicitly describe the result data set from complicated statistical processes.

Objectives

- Explicitly describe the result data set from complicated statistical processes.
- Promote the use of existing vocabularies.

Objectives

- Explicitly describe the result data set from complicated statistical processes.
- Promote the use of existing vocabularies.
- Provide metadata flexibility and extensibility.

- Explicitly describe the result data set from complicated statistical processes.
- Promote the use of existing vocabularies.
- Provide metadata flexibility and extensibility.
- Be flexible between encoding formats.

- Resource Description Framework
- WMO and Observable Property
- Flexible encoding formats
- UncertML
- Climate and Forecasting Conventions

Resource Description Framework

- Numerous authoritative sources are publishing vocabularies as RDF:
    - promoting reuse and inter-operation;

Resource Description Framework

- Numerous authoritative sources are publishing vocabularies as RDF:
  - promoting reuse and inter-operation;
- http://codes.wmo.int
- http://reference.metoffice.gov.uk
- RDF - Linked Data for flexible data formats.

WMO and Observable Property

Describing
Statistical
Process
Results

Mark Hedley -
Analysis,
Visualisation
and Data
Team - UK
Met Office

Objectives

Opportunities

Challenges

Examples

Conclusions

- http://wis.wmo.int/AvXML/AvXML-1.1/index.htm
- The XML encoding is targeted at WMO Manual on Codes 306 I.3.

Met Office

- http://wis.wmo.int/AvXML/AvXML-1.1/index.htm
- The XML encoding is targeted at WMO Manual on Codes 306 I.3.
- http://codes.wmo.int is publishing WMO codes table as RDF linked data.

Flexible encoding formats

Describing
Statistical
Process
Results

Mark Hedley -
Analysis,
Visualisation
and Data
Team - UK
Met Office

Objectives

Opportunities

Challenges

Examples

Conclusions

- netCDF
- HDF
- JSON
- XML

Flexible encoding formats

- netCDF
- HDF
- JSON
- XML
- Flexibility of content is valuable.
- Large numerical data sets mandate the use of binary encodings.

- http://www.uncertml.org/
- A mature vocabulary for defining statistical results.
- A large body of work has gone into this community standard, which we can benefit from.

Climate and Forecasting Conventions for netCDF files

Describing
Statistical
Process
Results

Mark Hedley -
Analysis,
Visualisation
and Data
Team - UK
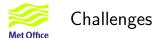Met Office

Objectives

Opportunities

Challenges

Examples

Conclusions

- http://cfconventions.org
- CF is valuable, but it suffers from inflexible metadata structures, particularly for:
    - standard names for phenomena;
    - cell methods for aggregations.

- Ordered Process chains and order neutral process chains.
- Namespaces, prefixes and identifiers.
- Compound and Nested metadata structures.
- I'm sure there are many others.

- Metarelate is hosting a collaboration space on Github for these topics:
- https://github.com/metarelate/statistical-process/issues

- RDF, Linked Data (LD), provides a flexible extensible approach to encoding metadata.

Conclusions

- RDF, Linked Data (LD), provides a flexible extensible approach to encoding metadata.
- We can use published vocabularies and adopt useful models and definitions from numerous domains.

Met Office

- RDF, Linked Data (LD), provides a flexible extensible approach to encoding metadata.
- We can use published vocabularies and adopt useful models and definitions from numerous domains.
- A standardised methodology for encoding RDF metadata in netCDF and HDF will provide a hugely valuable platform to work from.

Conclusions

Describing
Statistical
Process
Results

Mark Hedley -
Analysis,
Visualisation
and Data
Team - UK
Met Office

- RDF, Linked Data (LD), provides a flexible extensible approach to encoding metadata.
- We can use published vocabularies and adopt useful models and definitions from numerous domains.
- A standardised methodology for encoding RDF metadata in netCDF and HDF will provide a hugely valuable platform to work from.
- netCDF-LD
- HDF-LD

Met Office

Describing
Statistical
Process
Results

Mark Hedley -
Analysis,
Visualisation
and Data
Team - UK
Met Office

- Questions?
- Comments?
- Thoughts?