Characterisation of metadata to enable high quality climate applications and services

Deliverable D400.3
Concrete encodings of Commentary metadata

Partners providing input: Met Office, University of Reading, Infoterra, DWD, KNMI, SIH

version number	date	comments
0.2	29th October 2013	First draft for review by partners
1.0	2nd December 2013	First release

CHARMe is funded by the EC under its FP7 Research Programme

Document Control

Contributors

Person	Role	Organisation	Contribution
M. Nagni	Software Engineer	STFC	Main author

Document Approval

Person	Role	Organisation

List of Acronyms

CHARMe	CHARacterisation of Metadata
URI	Uniform Resource Identifier
OA	Open Annotation
FOAF	Friend Of A Friend
SRD	Software Requirements Document
TBD	To Be Defined
UC	Use Case
RDF	Resource Description Framework

References

ID	Author	Document Title	Date
[R-1]	P J Kershaw	Deliverable 400.1 Data Model for "Commentary Metadata"	28th October 2013
[R-2]	Project partners	Characterisation of metadata to enable high-quality climate applications and services, project proposal	23rd November 2011
[R-3]	B.N Lawrence, R Lowry, P Miller, H Snaith and A Woolf.	Information in environmental data grids, Phil. Trans. R. Soc. A vol. 367 no. 1890 1003-1014	13th March 2009
[R-4]	STFC	Analysis of Existing Technical Solutions	14th May 2013
[R-5]	University of Reading	User Requirements Document, version 1.2	21st October 2013
[R-6]	CGI	System Requirements Document	TBD

Revision History

Issue	Author	Date	Description
0.1	M. Nagni	24 October 2013	Initial Draft.
0.2	M. Nagni, P J Kershaw	29 October 2013	Draft for review by partners
1.0	M. Nagni, P J Kershaw	2 December 2013	First release

Contents

1 Introduction

The goal of this document is to describe a set of example encodings of C-metadata based on the guidelines defined in CHARMe Data Model [R-1]. The model identifies a number of resource types (datasets, dataset collections, etc.) and a supporting framework that applies a number of ontologies (OpenAnnotation[1], Cito[2], Foaf[3], etc.) in order to realize the C-Metadata concept.

The C-Metadata concept can be effectively modelled in RDF[4], where the simplest relation between two objects is expressed in the form of subject-predicate-object. The first part of this document describes different types of triples; later these are applied in a number of use cases each with increasing increments in complexity.

A number of different encodings for RDF are in common use including Turtle, RDF/XML and JSON-LD. Turtle provides a simple human-readable form. This has been used in the Data Model document and for correspondence within the project to communicate and debate concepts and ideas. For machine-machine interfaces, the preferred syntax to express RDF information will be JSON-LD[5]. JSON-LD is less verbose than XML, allows for easy integration with Javascript and has widespread support with other programming languages.

[1] http://www.openannotation.org/spec/core
[2] http://purl.org/spar/cito/
[3] http://www.foaf-project.org/docs/specs
[4] http://www.w3.org/TR/2004/REC-rdf-mt-20040210/
[5] http://www.w3.org/TR/json-ld/

2 Basic triples

A typical CHARMe annotation contains a rich set of information as who, when, why the annotation was created; in the following examples many details have been omitted for the sake of clarity in order to highlight the particular feature in question. One of the great features of RDF is that any triple can be updated easily at anytime. Many of the examples below use this feature in order to demonstrate how solve a specific problem without having to duplicate unnecessary information.

For sake of simplicity, because is not in the scope of this document to describe the CHARMe software architecture, we assume that all the examples are extracted from the triplestore of an organization called WellknownHost (WH).

2.1 Open Annotation

CHARMe C-Metadata is expressed as the linking of resources, OA body and target in an annotation. Each is represented by a URI. When a new annotation is made a URI must be generated and associated with it.

We will not enter into the details of how the client or the server generate a new URI, however note that the URI is useful not just to identify a resource by ID but also, following the Linked Data recommendations[6], it should be possible to resolve the resource and thus be able to expose it outside the triplestore on the world-wide web..

[6] http://www.w3.org/DesignIssues/LinkedData.html

2.1.1 Text Comment Annotation

This along with 2.1.2 is the simplest case. An annotation is created to associate a *comment*, the body, to a *resource*, the target.

```
{
 "@graph": [
        {
         "@id": "http://wellknownhost.org/kp_9870",
         "@type": [
               "http://www.w3.org/2011/content#ContentAsText",
               "http://purl.org/dc/dcmitype/Text"
               ],
          "http://purl.org/dc/elements/1.1/format": "text/plain",
          "http://www.w3.org/2011/content#chars": " Well resolved spatial data ",
          "http://www.w3.org/2011/content# characterEncoding ": " utf-8 "
         },
         {
          "@id": "http://wellknownhost.org/ kp_12345",
          "@type": "http://www.w3.org/ns/oa#Annotation",
          "http://www.w3.org/ns/oa#hasBody": {
            "@id": "http://wellknownhost.org/kp_9870"
           },
           "http://www.w3.org/ns/oa#hasTarget": {
            "@id": "http://data.example.org/thing/1"
           }
        }
    ]
 }
```

Figure 1: Text comment annotation

In this example a WH application created an OA (URI: kp_12345) to associate a *comment* (the body, URI: kp_9870) to the *resource* (the target, URI: http://data.example.org/thing/1).
Note that, while that the target *resource* is a previously existing object which already has an associated URI, the text of the *comment* is created in the same moment of the OA forcing the WH to generate two URI: one for the annotation and for the body.

2.1.2 Associating two resources

A user wants to express an association between two resources.

```
{
 "@graph": [
        {
         "@id": "http://wellknownhost.org/ kp_12345",
         "@type": "http://www.w3.org/ns/oa#Annotation",
         "http://www.w3.org/ns/oa#hasBody": {
           “@id": " http://store.external.ac.uk/exp/23"
         },
         "http://www.w3.org/ns/oa#hasTarget": {
           "@id": "http://data.example.org/thing/1"
         }
       }
    ]
 }
```

Figure 2: Associating two resources

2.1.3 Preparing an Annotation

A user wants to generate an annotation to mark a specific resource but to specify the body later.

```
{
 "@graph": [
        {
         "@id": "http://wellknownhost.org/kp_12345",
         "@type": "http://www.w3.org/ns/oa#Annotation",
          "http://www.w3.org/ns/oa#hasTarget": {
            "@id": "http://data.example.org/thing/1"
          }
        ]
 }
```

Figure 3: Preparing an Annotation

A Body may be added to the annotation later, for example, explaining the importance of a dataset target and thus why it was bookmarked.

2.1.4 Identify the Information Type

Given a resource URI in principle it is possible to retrieve its data type. However there are several reasons why WH would like to specify anyway, inside the annotation, the data type of a resource:

- Despite is possible to resolve the URI to a URL, it may or may not contain type information
- It avoids the need for WH to resolve an external URI
- It allows WH to apply a custom classification

The previous considerations may be associates either to a 'body' resource or a 'target' resource or both. Using the 2.1.1 example is possible to solve the problem, for the target, adding the following triple

```
{
  "@graph": [
    {
      "@id": "http://data.example.org/thing/1",
      "@type": "http://purl.org/dc/dcmitype/Dataset"
    }
 ]
}
```

Figure 4: Identify the Information Type

In this case, we identify the resource < http://data.example.org/thing/1> as a dctype:Dataset. Please note that this triple:

- may be inserted also in a second time;
- is not specific to the annotation, that is, it defines the resource as a dctype:Dataset globally across the triplestore and if visible to the outside, to external agents

2.1.5 Providing a *Motivation* for the annotation

An annotation associates two resources: a “Target” and a “Body”. The example 2.1.3 demonstrates how updating the information associated with a resource impacts not only the annotation which includes it but the overall triple store connections. For this reason OA provides for **oa:motivatedBy** property for the annotation itself. For further explanations please refer to the [R-1, 1.4.1] or to the OA specifications.
For example, is possible to update the example 2.1.1 with

```
{
 "@graph": [
  {
   "@id": "http://wellknownhost.org/kp_12345",
   "http://www.w3.org/ns/oa#motivatedBy": {
    "@id": "http://www.w3.org/ns/oa#bookmarking"
   }
  }
 ]
}
```

Figure 5: Providing a Motivation for the annotation

2.1.5.1 Tagging a resource

A user wants to associate a label with a resource, i.e. an image. Then the annotation would look like this:

```
{
 "@graph": [
        {
         "@id": "http://wellknownhost.org/kp_9870",
         "@type": [
               "http://www.w3.org/2011/content#ContentAsText",
               "http://purl.org/dc/dcmitype/Text"
               ],
          "@type": [
               " http://www.w3.org/ns/oa#Tag"
                 ],
         "http://purl.org/dc/elements/1.1/format": "text/plain",
         "http://www.w3.org/2011/content#chars": "Oxford Airport area",
         "http://www.w3.org/2011/content#characterEncoding ": "utf-8 "
         },
         {
          "@id": "http://wellknownhost.org/kp_12345",
          "@type": "http://www.w3.org/ns/oa#Annotation",
          "http://www.w3.org/ns/oa#hasBody": {
                       "@id": "http://wellknownhost.org/kp_9870"
           },
           "http://www.w3.org/ns/oa#hasTarget": {
                        "@id": "http://data.example.org/thing/23"
           },
            "http://www.w3.org/ns/oa#motivatedBy": {
                       "@id": "http://www.w3.org/ns/oa#tagging"
            }
          },
          {
           "@id": "http://data.example.org/thing/23",
           "@type": "http://purl.org/dc/dcmitype/Image"
           "http://purl.org/dc/elements/1.1/format": "image/jpeg"
          }
        ]
 }
```

Figure 6: Associate a label to a resource

Note that now the oa:motivatedBy is set as oa:tagging.

2.1.5.2 Using semantic tagging

A Tag is a label attached to someone or something for identification or other information. Typically tags are created attaching a free text label to a resource. With the emergence of the Linked Open Data method, semantic tagging or tagging through URIs is gaining popularity.

```
{
 "@graph": [
  {
   "@id": "http://wellknownhost.org/f99f3991608a4d8185d2c0ff5e3247ce",
   "@type": "http://www.openannotation.org/spec/core/SemanticTag",
   "http://xmlns.com/foaf/0.1/page": {
    "@id": "http://vocab.ndg.nerc.ac.uk/term/P220/1/26"
   }
  },
  {
   "@id": " http://wellknownhost.org/c3f75541f61f439c9f7bdde08e924d34",
   "@type": "http://www.openannotation.org/spec/core/Annotation",
   "http://www.openannotation.org/spec/core/hasBody": {
    "@id": " http://wellknownhost.org/f99f3991608a4d8185d2c0ff5e3247ce"
   },
   "http://www.openannotation.org/spec/core/hasTarget": {
    "@id": "http://badc.nerc.ac.uk/view/badc.nerc.ac.uk__ATOM__dataent_ECMWF-OP"
   },
   "http://www.openannotation.org/spec/core/motivatedBy": {
    "@id": "http://www.openannotation.org/spec/core/tagging"
   }
  },
  {
   "@id": "http://badc.nerc.ac.uk/view/badc.nerc.ac.uk__ATOM__dataent_ECMWF-OP",
   "@type": "http://purl.org/spar/fabio/MetadataDocument"
  }
 ]
}
```

Figure 7: Using semantic tagging

2.1.6 Specify the annotation provenance

Knowing who, when, how created the annotation is critical to CHARMe as well to WellknownHost.

2.1.6.1 Specify the annotation creation time

```
{
  "@graph": [
    {
      "@id": "http://wellknownhost.org/kp_12345",
      "http://www.w3.org/ns/oa#annotatedAt": "2013-05-28T12:00:00Z"
    }
  ]
}
```

Figure 8: Specify the annotation creation time

2.1.6.2 Specify the annotation first serialisation time and subsequent updates

As anything else, the annotations may undergo revisions. The OA defines a property called **oa:serializedAt** and declares "*The annotation graph MUST have changed for this property to be updated, and as such represents the last modified datestamp for the Annotation.*" . This mean that the example 2.1.3 should not update this property on the other hand the example 2.1.4 shall do it. Updating this property is a simple as below

```
{
  "@graph": [
    {
      "@id": "http://wellknownhost.org/kp_12345",
      "http://www.w3.org/ns/oa#serializedAt": "2013-12-28T12:00:00Z"
    }
  ]
}
```

Figure 9: Specify the annotation first serialisation time and subsequent updates

2.1.6.3 Specify who created the annotation

For this goal, while the OA recommends the use of a foaf:Person class, the document [R-1, 1.4.3.1] requires it and specifies which properties should be associated with it.
Creation of the account and creation of the annotation are two independent steps consequently we can imagine two separate submission
Create the account:

```
{
 "@graph": [
  {
    "@id": "http://wellknownhost.org /kp_xs02300",
    "@type": "http://xmlns.com/foaf/0.1/Person",
    "http://xmlns.com/foaf/0.1/mbox": {
      "@id": mailto:maurizio.nagni@example.org   .
    },
    "http://xmlns.com/foaf/0.1/name": "Maurizio Nagni"
  }
 ]
}
```

Figure 10: Create a user account

Nb. in the above, specifying mbox is OPTIONAL. See [R-1, 1.4.1.3]

Incorporating the author information into the annotation (extends the example 2.1.1):

```
{
 "@graph": [
  {
    "@id": "http://wellknownhost.org/kp_12345",
   "http://www.w3.org/ns/oa#annotatedBy": {
          "@id": "http://proteus.badc.rl.ac.uk:8000/kp_xs02300"
    }  }
 ]
}
```

Figure 11: Associate a user to an annotation

2.1.7 Citation of a Dataset

The example in 2.1.3 showed cases where we want to save locally information that could be retrieved anyway resolving the URI. A specific case may arise when either the *target* or the *body*, or both, are of a document type and in this case CHARMe preferred solution is to use the CITO ontology


```
{
 "@graph": [
  {
   "@id": "http://badc.nerc.ac.uk/view/badc.nerc.ac.uk__ATOM__ACTIVITY_df1f7676-df0a-11e2-9431-
00163e251233",
   "@type": "http://purl.org/spar/fabio/MetadataDocument"
  },
  {
   "@id": "http://wellknownhost.org /44eb6ea67a2745448b73d4e8dfc1414f",
   "@type": "http://purl.org/spar/cito/CitationAct",
   "http://purl.org/spar/cito/hasCitationEvent": {
    "@id": "http://purl.org/spar/cito/citesAsDataSource"
   },
   "http://purl.org/spar/cito/hasCitedEntity": {
    "@id": "http://badc.nerc.ac.uk/view/badc.nerc.ac.uk__ATOM__ACTIVITY_df1f7676-df0a-11e2-9431-
00163e251233"
   },
   "http://purl.org/spar/cito/hasCitingEntity": {
    "@id": "http://dx.doi.org/10.5285/4BCFA3A4-C7EC-4414-863D-CAECEB21F16F"
   }
  },
  {
   "@id": "http://wellknownhost.org /bbfe03ce91754db09b89299b8e766f90",
   "@type": "http://www.openannotation.org/spec/core/Annotation",
   "http://www.openannotation.org/spec/core/hasBody": {
    "@id": "http://wellknownhost.org/44eb6ea67a2745448b73d4e8dfc1414f"
   },
   "http://www.openannotation.org/spec/core/hasTarget": {
    "@id": "http://badc.nerc.ac.uk/view/badc.nerc.ac.uk__ATOM__ACTIVITY_df1f7676-df0a-11e2-9431-
00163e251233"
   },
   "http://www.openannotation.org/spec/core/motivatedBy": {
    "@id": "http://www.openannotation.org/spec/core/linking"
   }
  },
  {
   "@id": "http://dx.doi.org/10.5285/4BCFA3A4-C7EC-4414-863D-CAECEB21F16F",
   "@type": "http://purl.org/spar/fabio/ConferencePaper"
  }
 ]
}
```


Figure 12: Citation of a Dataset

2.1.8 Full Example

Finally we can collect all our examples and assemble a full annotation for citing a dataset.


```
{
 "@graph": [
  {
   "@id": "http://dx.doi.org/10.5285/4BCFA3A4-C7EC-4414-863D-CAECEB21F16F",
   "@type": "http://purl.org/spar/fabio/ConferencePaper"
  },
  {
   "@id": "http://wellknownhost.org /kp_xs02300",
   "@type": "http://xmlns.com/foaf/0.1/Person",
   "http://xmlns.com/foaf/0.1/mbox": {
    "@id": "mailto:maurizio.nagni@example.org"
   },
   "http://xmlns.com/foaf/0.1/name": "Maurizio Nagni"
  },
  {
   "@id": "http://wellknownhost.org /ad9671845715422bb768b85e15335749",
   "@type": "http://purl.org/spar/cito/CitationAct",
   "http://purl.org/spar/cito/hasCitationEvent": {
    "@id": "http://purl.org/spar/cito/citesAsDataSource"
   },
   "http://purl.org/spar/cito/hasCitedEntity": {
    "@id": "http://badc.nerc.ac.uk/view/badc.nerc.ac.uk__ATOM__ACTIVITY_df1f7676-df0a-11e2-9431-00163e251233"
   },
   "http://purl.org/spar/cito/hasCitingEntity": {
    "@id": "http://dx.doi.org/10.5285/4BCFA3A4-C7EC-4414-863D-CAECEB21F16F"
   }
  },
  {
   "@id": "http://wellknownhost.org /dce2ebe614674707a21a67b7fd430b3c",
   "@type": "http://www.openannotation.org/spec/core/Annotation",
   "http://www.openannotation.org/spec/core/annotatedAt": "2013-05-28T12:00:00Z",
   "http://www.openannotation.org/spec/core/annotatedBy": {
    "@id": "http://proteus.badc.rl.ac.uk:8000/kp_xs02300"
   },
   "http://www.openannotation.org/spec/core/hasBody": {
    "@id": "http://wellknownhost.org /ad9671845715422bb768b85e15335749"
   },
   "http://www.openannotation.org/spec/core/hasTarget": {
    "@id": "http://badc.nerc.ac.uk/view/badc.nerc.ac.uk__ATOM__ACTIVITY_df1f7676-df0a-11e2-9431-00163e251233"
   },
   "http://www.openannotation.org/spec/core/motivatedBy": {
    "@id": "http://www.openannotation.org/spec/core/linking"
   },
   "http://www.openannotation.org/spec/core/serializedAt": "2013-12-28T12:00:00Z"
  },
  {
   "@id": "http://badc.nerc.ac.uk/view/badc.nerc.ac.uk__ATOM__ACTIVITY_df1f7676-df0a-11e2-9431-00163e251233",
   "@type": "http://purl.org/spar/fabio/MetadataDocument"
  }
 ]
}
```


Figure 13: Full example for citing a dataset