

Necessities on a Descriptive Level for Reusing Metadata Descriptions

OHYA Kazushi

Maruzen Co.,Ltd.

Nihombashi, Chuo-ku, Tokyo 103-8245, Japan

ohya@maruzen.co.jp

Abstract

The present RDF is opaque in reusing the descriptions in ready-made metadata repositories. To enhance reusability of descriptions in RDF/XML, we need a unit for packaging a set of descriptions on a semantic schema level. And, we may need a new rule to handle a composite unit of not linear-fold but link-based or graph data structure, and serialize it to reuse.

Keywords: RDF, reuse, packages, composites

1 Introduction

This paper is a proposal from metadata communities for requirements to enhance reusability of metadata descriptions. The present RDF specification is opaque in using a particular metadata description as a base for another metadata description. In the following section, first, we observe the way of reusing descriptions in HTML, XML, and RDF(S). Next, we propose to introduce a new element of DC as a unit of reuse. Last, we consider a common way to enhance reusability.

2 Reuse

Reusing of descriptions is realized as partial sharing and localization of existing descriptions.

2.1 Partial sharing

To handle partial data units¹, link features are used in SGML/XML. For example, XLink provides facilities to indicate partial units in SGML/XML instances with some conditions.

```
<any xlink:type="simple"
      xlink:href="#xpointer(id('item1'))"/>
```

Example 1

Reference-Referent relationships are represented as link features extended by XLink. Addressing of partial units is supported by a fragment identifier # in URL, and conditions defined by XPointer.

Link features can be used not only to indicate objects but also to relate with some processes. The resultant descriptions appear after indicating

location only, or with embedding processes. We say partial sharing in both of the cases.

2.2 Particularization

In reusing existing instances, in addition to the partial sharing, or referring and copying, it is useful to modify a part of the instances. This aspect can be expressed as particularization, localization, and, in term of OO ways, instantiation.

```
<dc:creator>OHYA Kazushi</dc:creator>
<any xlink:type="simple" xlink:show="embed"
      xlink:href="#xpointer(id('item1'))/dc:title"/>
<dc:date>2001-06-20</dc:date>
```

Example 2

The example 2 is a complete partial sharing of ready-made metadata. In the following, particular descriptions are modified.

```
<any xlink:type="simple" xlink:show="embed"
      xlink:href="#xpointer(id('item1'))">
  <dc:date>2001-06-20</dc:date>
  <dc:description>
    <localRequestNumber>MBN0100001
  </localRequestNumber></dc:description>
</any>
```

Example 3

The example 3 is intended to use linked documents as a mold, and override the part of `dc:date` and `dc:description`².

3 Reusing in the present specifications

In the following, we observe how to realize the partial sharing and the particularization in the present specifications.

3.1 In HTML

Essentially HTML provides no facility for reusing descriptions.

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 3.2 Final//EN">
<html>
  <head>
    <meta name="dc.creator" content="TOUKYOU Taro">
    <meta name="dc.date" content="2000-01-01">
  </head>
  <body><h2>This is Taro's web page.</h2>
  <p><meta name="dc.date" content="2001-01-15">
  This is Ohya's comment</p></body></html>
```

Example 4

¹It is expressed "sub-resources" in XPointer[4].

²However, it is not undefined in the present specifications. And this will be a point of our proposal.

This is an ordinary example of metadata descriptions in HTML documents.

For external documents, this kind of embedded metadata is not suitable for referents, because there is no mechanism to refer to just head elements in HTML.

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0//EN"
<html><head>
<link href="http://kazushi.ohya.civ.jp/dclist.xml"
      title="metadata" rel="DC lists" ></head>
<base href="http://kazushi.ohya.civ.jp/dclist.xml">
<body>
<metadata>
<iframe type="text/plain" src="#item1" frameborder="0"
        scrolling="no" height="" height="">
</metadata></body></html>
```

Example 5

The example 5 is a forced description. It is intended to indicate an external metadata file by a link element in a head element, and try to embed metadata by an element `iframe`. However, substantially, the element `link` is not available on the original purpose. And, `iframe` is originally not for embedding descriptions into raw data structures, but for setting other dimension for formatted HTML documents.

3.2 In XML

In XML many facilities are provided to indicate partial descriptions and the metadata.

3.2.1 Common XML

The example 6 is a list of bibliography, which is referred from another instance named `ohya1.xml` in the example 7.

```
<?xml version="1.0" ?>
<!-- XML Instance
      named http://kazushi.ohya.civ.jp/mybiblist1.xml -->
<mylist>
  <item id="item1">
    <title lang="en" notation="">Asleep</title>
    <title lang="ja" notation="kunrei">utataka</title>
    <author>YOSHIMOTO Banana</author></item>
  <item id="item2">
    <title lang="ja" notation="kunrei">totechiteta</title>
    <author>NOSAKA Akiyuki</author></item></item>
<!-- contd. -->
</mylist>
```

Example 6

```
<?xml version="1.0" ?>
<!-- XML Instance
      named http://kazushi.ohya.civ.jp/ohya1.xml -->
<myFavoritBook
  xmlbase="http://kazushi.ohya.civ.jp/mybiblist1.xml">
  <item id="mfbi">
    <include xlink:type="simple" xlink:show="embed"
      xlink:href="#xpointer(id('item1'))/*"/>
    <comment><!--contd.--></comment></item></myFavoritBook>
```

Example 7

In the example 7, we use an element `include` with XLink features. The actual description is

³This mechanism may require a change in traditional SGML/XML serialization rule, and some descriptive rules[10]. We will treat more detail later.

partially implanted from the external instance `mybiblist1.xml`. We succeed in partial sharing in XML. The following is an example of particularization.

```
<?xml version="1.0" ?>
<!-- XML Instance
      named http://kazushi.ohya.civ.jp/ohya2.xml -->
<myFavoritBook
  xmlbase="http://kazushi.ohya.civ.jp/mybiblist1.xml">
  <item id="mfbi">
    <include xlink:type="simple" xlink:show="embed"
      xlink:href="#xpointer(id('item2'))/author">
    <title lang="ja" notation="hepburn">totechiteta</title>
  </include><comment><!-- contd. --></comment>
</item></myFavoritBook>
```

Example 8

This example is intended to rewrite the content of `title` in different notation. If it is possible to import a whole part of `item2` and override the `title` part with a new description, it will reduce writing cost and be very useful. However, it is impossible in the present specifications³.

3.2.2 Common RDF

In the present SGML/XML style, metadata will be described in RDF.

```
<?xml version="1.0" ?>
<!-- XML Instance
      named http://kazushi.ohya.civ.jp/mybib.rdf.xml -->
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description ID="dcitem1">
    <dc:title xml:lang="ja">totechiteta</dc:title>
    <dc:creator>NOSAKA Akiyuki</dc:creator>
  </rdf:Description>
  <rdf:Description ID="dcitem2">
    <dc:title xml:lang="en">Asleep</title>
    <dc:title xml:lang="ja">utataka</title>
    <dc:creator>YOSHIMOTO Banana</dc:creator>
  </rdf:Description>
</rdf:RDF>
```

Example 9

The following is an example of partial sharing.

```
<?xml version="1.0" ?>
<!-- XML Instance
      named http://kazushi.ohya.civ.jp/ohya3.xml -->
<top>
  <rdf:RDF
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:dc="http://purl.org/dc/elements/1.1/">
    <rdf:Description about="#book1">
      <include xlink:type="simple" xlink:show="embed"
        xlink:href="http://kazushi.ohya.civ.jp/mybib.rdf.xml
          #xpointer(id('dcitem1'))/*"/>
    </rdf:Description>
    <rdf:Description about="#book2">
      <dc:title
        rdf:resource="http://kazushi.ohya.civ.jp/mybib.rdf.xml
          #xpointer(id('dcitem2'))/title[@xml:lang='ja']"/>
      <dc:creator
        rdf:resource="http://kazushi.ohya.civ.jp/mybib.rdf.xml
          #xpointer(id('dcitem2'))/dc:creator"/>
    </rdf:Description>
    </rdf:RDF>
  <body><book name="book1"><!-- contd. --></book>
    <book name="book2"><!-- contd. --></book>
  </body></top>
```

Example 10

In the example 10, the partial descriptions are referred in two ways. In the first `rdf:Description`, XLink features are used in an element include. Since RDF has no function to set link behavior, importing partial descriptions is indicated by XLink features. In the second, an attribute `rdf:resource` is used in a `dc` element. In the present RDF, `ID`, `about`, and `resource` are alternative referential attributes. Then, if we reuse ready-made descriptions in RDF, we must set referred targets in every possible elements in advance.

For particularization, it is useful for the above first `rdf:Description` to be described as follows.

```
<rdf:Description about="#book1">
  <set id="ohya3:book1"
    xlink:type="simple" xlink:show="embed"
    xlink:href="http://kazushi.ohya.civ.jp/mybib.rdf.xml
    #xpointer(id('dcitem1.set1'))">
    <dc:title xml:lang="ja">totechiteta</dc:title>
  </set></rdf:Description>
```

Example 11

The description is intended to override the element `dc:title` with the value `ja` in the attribute `lang`⁴. As above, in order to set a scope for overridden resources, we use a new element `set` here. An attribute `resource` cannot be used for this purpose, because `resource` can appear only in empty elements, so we cannot handle sub-elements of `rdf:Description` with an attribute `resource`.

3.2.3 XML Namespace in RDF

RDF is a metalanguage not only for semantic definitions but also common descriptions to manifest Meta-Object relationship on a descriptive level. Then, for example, RDF can be used as a metalanguage for XML Namespace definitions. It is very beneficial to use XML Namespace resources as ready-made repositories. XML Namespace can be used as a system of ready-made description repositories because it makes possible to have structural sub-elements. The following example 12 is a localized partial description of an ISBN Namespace resource.

```
<top xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:isbn="http://isbn.org/">
  <isbn:4-621-04890-2>
    <localarea>
      <ndc>060</ndc>
      <requestNumber>111</requestNumber></localarea>
    </isbn:4-621-04890-2></top>
```

Example 12

The example 12 can be described with RDF as follows.

```
<top xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
```

⁴The resultant is changed in notation from kunrei to hepburn in Japanese.

⁵DCMI has provided a set of semantic elements. Then, I will not represent syntactic definitions here.

```
xmlns:isbn="http://isbn.org/">
  <rdf:Description about="#book3">
    <isbn:4-621-04890-2 ID="local123">
      <localarea>
        <ndc>060</ndc>
        <requestNumber>111</requestNumber></localarea>
      </isbn:4-621-04890-2></rdf:Description></rdf:RDF>
<!-- contd. --></top>
```

Example 13

In the example 13, `rdf:Description` is used as a unit packaging metadata. However, since referential attributes in RDF are alternative, we need other units for overriding the sub-elements. The element `isbn:*` works as such a unit. Setting a referred unit not on a syntactical level like defined element sets of RDF but on semantic schema like DC is valuable for overriding and reusing ready-made descriptions.

4 Proposal

To make RDF possible to be reused, we propose to set a new unit for a composite of Dublin Core elements. It is just like a package in the Warwick Framework. For convenience, I use `package` as the name. The package is defined as that $[[\text{PACKAGE}]] = \{x | x \text{ is a set of DC elements.}\}$ ⁵. The expected descriptions are as follows.

```
<dc:package id="item1">
  <dc:title>Language</dc:title>
  <dc:creator>Edward Sapir</dc:creator>
  <dc:date>1921</dc:date>
</dc:package>
```

Example 14

This usage requires the present SGML/XML to be revised in syntactic or descriptive rules.

4.1 Descriptive rules

To make packaging units improved in usability with a way of overriding specific descriptions like the example 8, 11, and 13, we need to set a rule of abbreviation in descriptions. For example, it includes a rule of a filling-in direction like inheritance and propagation, priority, and so on.

As such a rule concerning descriptions, I presuppose the four stage; **syntactic rules**, **descriptive rules**, **set phrases**, and **descriptive styles**. These compose a hierarchy in the restrictive strength. The syntactic rule defines syntax of metalanguages like SGML/XML specifications. The descriptive rule restricts the descriptive power metalanguages have originally. Abbreviation rules are defined on this level. The set phrase is a preserved set of elements. RDF is one of the set to manifest a Meta-Object relationship on a descriptive level. The last descriptive style includes common writing rules, e.g. cataloging rules, orthography and so on. Descriptive rule studies are

important to enhance reusability and persistency of SGML/XML descriptions.

5 Reusability with persistency

To be persist, descriptions need to be pure as the type and independent from the others. SGML/XML has deficiencies in this point.

5.1 Implicit and Explicit modeling

As a way to set a composite object there are implicit and explicit approaches[1].

In implicit modeling, related descriptions are embedded into base descriptions like the example 4. The resultant descriptions are unsuitable for reuse in that they are dependent on the context. In SGML/XML, explicit models are described with links in base descriptions like the above example 7 and 10. However, this way ensures independency of related descriptions, but not the base descriptions themselves, because they contain link features specific to related ones. In essence, base descriptions should be more pure than related descriptions like metadata, style definitions, schema, and so on. Then, we need a mechanism to ensure a relationship between base and related, or SGML/XML instances. It is very similar to a function of RDF in that RDF manifests a relationship between descriptions.

5.2 Sharing objects in link-based models

In SGML/XML, a composite unit linked to component descriptions is called a hub-document/instance. However, a hub-instance has a serious problem of a structural collision. There is no mechanism to mediate multiple data structures in SGML/XML.

Then, we presuppose that we may had better divide data models between linear-fold and link-based or graph data structures, and prepare mechanism to handle each model. In the present specifications, mechanisms to handle a link-based data model are devised for a single instance, e.g. Grove and DOM. To handle multiple link-based instances, we need especially a rule for serializing objects in link-based models to linear-fold descriptions⁶. RDF can be used for the way to make them clear, and it is worth being verified. Reusability of the descriptions in RDF is important in this scene.

6 Conclusion

In this paper, we propose to introduce a new unit for packaging a set of description on a schema level to enhance reusability of the descriptions. And, we consider necessities for reusing descriptions. This paper is in a series of our research on descriptive rules in SGML/XML.

References

- [1] A.Artale, E.Franconi, N.Guarino, and L.Pazzi, 1996, "Part-Whole Relations in Object-Centered Systems: An Overview", *Data & Knowledge Engineering*, Elsevier
- [2] P.Champin, 2001, "RDF Tutorial", online at <http://www710.univ-lyon1.fr/~champin/>
- [3] M.Dahchour, M.Kolp, A.Pirotte, and E.Zimányi, 1998, *Generic Relationships in Information Modeling*, Technical Report IAG-QANT, University of Louvain
- [4] S.DeRose, E.Maler, and R.Daniel Jr., 2001, "XML Pointer Language (XPointer) version 1.0", WD-xptr-20010108, W3C
- [5] D.Maluf and G.Wielerhold, 1997, "Abstraction of Representation for Interoperation", *Lecture Notes in AI*, subseries of LNCS, Vol.1315
- [6] S.Melnik and S.Decker, 2000, "A Layered Approach to Information Modelig and Ineroperability on the Web", *Proceedings of ECDL'00 Workshop on the semantic Web*
- [7] S.Melnik, 2000, "Declarative mediation in distributed systems", *Proceedings of International Conference on Conceptual Modeling(ER'00)*
- [8] E.Miller, P.Miller, and D.Brickley, 1999, "Guidance on expressing the Dublin Core within the Resource Description Framework(RDF)", WD-dc-rdf-19990701, DCMI
- [9] K.Ohya and S.Tutiya, 2000, "A proposal to use text-based data models for pre-processes of making databases and data preservation (in Japanese)", *Proceedings of the 2nd Forum on Art Documentation*, Japan Art Documentation Society, and *Journal of Information Processing and Management*(JOHO KANRI in Japanese) Vol.43 No.2, Japan Science and Technology Corporation
- [10] K.Ohya and S.Tutiya, 2000, "Descriptive rules of SGML/XML for data preservation", *Proceedings of the 3rd International Conference of Asian Digital Library*
- [11] G.Wiederhold, 1994, "Interoperation, Mediation, and Ontologies", *Ptoceedings International Symposium on Fifth Generation Computer Systems(FGCS94)*

⁶Conversion form linear-fold to link-based objects may be supported with XML Namespace.