

Deriving the Properties of Object Types for Research Data Relation Model

Suntae Kim*

Korea Institute of Science and Technology Information
245 Daehak-ro, Yuseong-gu, Daejeon, Korea
E-mail: stkim@kisti.re.kr

ABSTRACT

In this study, the properties of the object types required to describe the relationship among research data resources, which may be generated during the life cycle of the research, are derived. The properties of Fedora Commons and DSpace, which are open source software used for resource management, and schema properties published in DataCite were analyzed. Based on relation names of Fedora Commons, nine new relation names were derived. Thirty-eight object type properties consolidating the target properties of the analysis were derived. The result of this study can be used as basic material for crosswalk research studies of object type relation terms to ensure interoperability among the systems.

Keywords: Object type relation, Fedora Commons, DSpace, DataCite, Research Data, Scientific Data

1. INTRODUCTION

High-performance observation, measurement, and laboratory equipment are utilized in various domains. The development of high-speed networks and the

related research environment has resulted in the generation of various types of data in large quantities. Thus it has become a responsibility for organizations which manage and service research data to build an environment for systematic data management and

Open Access

Accepted date: June 12, 2013

Received date: May 6, 2013

*Corresponding Author: Suntae Kim
Senior Researcher
Korea Institute of Science and Technology Information
245 Daehak-ro, Yuseong-gu, Daejeon, Korea
E-mail: stkim@kisti.re.kr

All JISTaP content is Open Access, meaning it is accessible online to everyone, without fee and authors' permission. All JISTaP content is published and distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0/>). Under this license, authors reserve the copyright for their content; however, they permit anyone to unrestrictedly use, distribute, and reproduce the content in any medium as far as the original authors and source are cited. For any reuse, redistribution, or reproduction of a work, users must clarify the license terms under which the work was produced.

data reuse. This is so because locating a desired data set and searching for relevant data among an enormous amount of data accurately and quickly will guarantee high productivity of their research projects.

This problem requires related ontology research. Soldatova and King (2006) emphasized early on that formal description of experiments for efficient analysis, annotation, and sharing of the results is a fundamental part of the practice of science and that ontology is required to achieve this objective. The properties required for describing resources may vary with the domain, but there may be a number of common properties that describe the relationship among the resources.

In this study, the object type properties that may be used for describing the relationship among the resources are examined and analyzed. In this process, integrated object type properties are derived and the related properties are categorized into groups.

2. RELATED WORK AND GOALS

Research and experimental scenarios include capturing data with the equipment, creating data through simulations, transferring data to the computer, performing computation, and storing the resulting data. Research involves various work steps. Building up the relationships among the data at each work step is necessary for searching and discovering research data and for reproducing the research later. Therefore, research studies on ontology that describe the relationship among data, i.e., the relationship among the resources, are required. However, so far, the ontology has focused on studying data type properties. Soldatova and King (2006) proposed the ontology EXPO and linked the SUMO (the Suggested Upper Merged Ontology) with a subject-specific ontology for experiments. Washington and Lewis (2008) claimed that ontology has made it easy to share scientific data. They conducted research on ontology usability for genetic studies. Especially, they emphasized that fast search and comparison of various data sources using the ontology would be useful. They claimed that there

would be no shortage of genetic data and pointed out that it would be rather a matter of how quickly one could obtain and analyze the desired data from a large amount of genetic data.

Goranova et al. (2011) suggested an ontology to acquire the meaning of general scientific data produced through observations and measurements. The proposed ontology can be used to describe the meaning of the scientific data generated through simulations and experiments in the field of physics. Shotton (2011) suggested an ontology by mapping the DataCite elements to the global ontology terms.

The above studies tend to focus on the data type properties used for describing the properties of resources in certain domains, but the studies focusing on describing the relationship among resources are still scanty.

3. RESEARCH METHODOLOGY

The Korea Institute of Science and Technology Information (KISTI) is developing the research data platform named P-CUBE.¹ The P-CUBE's major modules use open source software platforms such as Fedora Commons, DSpace, and MySQL. They are well known as robust open source software. KISTI is developing functions by using these open source software platforms which are recognized as powerful repository tools. Fedora Commons is used for a storage layer and DSpace is used for an application layer. Thus object type properties which are used in these systems need to be analyzed. DataCite announced a metadata schema to publish the data. In this schema, there are object type properties. The P-CUBE has the function to publish data by using the DataCite metadata. So the DataCite metadata Schema should be analyzed also.

As data properties vary with data types and research areas, property analysis was conducted, limited to the object type properties that describe relationships among the resources. Ontology terms declared in Fedora Commons (hereinafter referred to as "FCO"), Dublin Core - Library Application Profile² terms used in Dspace (hereinafter referred to as "DSO"), and the

¹The P-CUBE (Platform for Convergence and Unification of Big E-resources) is a platform providing easy access for safe storage and reuse of scientific data. (See <http://www.datacite.kr>)

²DC-Lib Application Profile (2013, April 3). DC-Lib Application Profile. Retrieved from <http://www.dublincore.org/documents/library-application-profile/>

terms defined in DataCite metadata schema³ standard (hereinafter referred to as “DCO”) were analyzed. The terms were grouped according to their uses. DSO and DCO terms were analyzed based on group names suggested by the Fedora Commons. The relation group level in the FCO was expanded to include properties declared in DCO and the DSO.

In addition to FCO, DSO, and DCO, a variety of ontology properties may exist to describe relationships among the resources. However, in this study, only the properties of Fedora Commons and DSpace, the resource management open source software used worldwide, were analyzed. Also, resource-relation properties of DataCite, an international community to provide a permanent approach to resources by allocating Digital Object Identifiers (DOI) to resources, were analyzed as well. Accordingly, this study is considered as a meaningful work of research analysis.

4. DSPACE RESOURCE RELATION MODEL ANALYSIS

The Resources managed in DSpace are used for the DSpace data management model. Community, the highest concept in DSpace, may include one or more sub-communities or collections. The collection that binds logically related items may also include one or more items. An item includes metadata and bitstreams. A bitstream means files collected in DSpace, and one may have relationships with other items, collections, and communities. It may have relationships with other bitstreams as well. In order to describe these resource relations, DSpace provides the following object type properties.

DSpace provides qualified Dublin Core (DC) as a default option. However, it is possible to set multiple schemas and to select metadata fields from the combination of these schemas. The item can have different types of descriptive metadata as bitstreams. Simple descriptive metadata for the communities and collections are stored in DBMS.

DSpace, which provides qualified DC as the default

schema, can use resource relation elements declared in DC.

<relation> elements declared in the DC metadata element set are used to describe the related resources, and the qualified DC declares the following 10 elements as their sub-elements:

- isVersionOf: This means a substantial change in the content, rather than a change of the format.
- isFormatOf: A resource with the same content as the related resource but expressed in a different format.
- hasFormat: A resource with another format related to it.
- isReplacedBy: Resource used, replaced, or discarded by the related resource.
- Replaces: Uses, replaces, or discards the current resource, instead of the related resource.
- isPartOf: A resource which is a physical or logical part of the related resource.
- hasPart: A resource which physically or logically includes the related resource.
- Requires: The current resource requires the related resource to support its function, delivery, or integrity.
- isReferencedBy: A resource physically or logically referenced by the related resource
- References: The current resource may refer to or cite the related resource, or point it out in a different way.

It also has the following two unique aspects:

First, it does not declare the <hasVersion> element. With regards to this, ‘DCMI-Libraries Working Group’ states that DSpace does not include <hasVersion> as the <isVersionOf> element can deliver the meaning more clearly than <hasVersion> does. In other words, the declaration ‘A isVersionOf B’ more clearly expresses that Version B was created before Version A,’ than the declaration ‘A hasVersion B’ does.

Second, it does not declare the <isRequiredBy> element. This element is declared in the qualified DC to describe the resource information physically or logically required by the related resources, but is not used in the DC-Lib Application Profile.

³DataCite Metadata Schema Repository.(2013, May 1). DataCite Metadata Schema RepositoryRetrieved from <http://schema.datacite.org/>

5. FEDORA COMMONS RESOURCE RELATION MODEL ANALYSIS

The resource relation model ontology provided by Fedora Commons consists of 22 object type properties, which can be divided into nine relation groups. It declares <fedoraRelationship> as the highest property. The relationship among the properties declared in the FCO is illustrated in Fig. 1. The FCO declares the properties to describe such relationships as Derivation, Equivalence, Dependency, Descriptive, Commentary, Metadata, Part/Whole, Membership, and Set Membership. The properties of 'Descriptive' relationship are declared as higher properties of 'Commentary' and 'Metadata' relationships, and the properties of 'Part/Whole' relationship as higher than 'Membership' and 'Set Membership' relationships.

6. DATACITE RESOURCE RELATION MODEL ANALYSIS

As of March 2013, DataCite version 2.2⁴, published in July 2011, is the latest version registered, and the metadata work group operated by DataCite is working on version 2.3.

A <relationType> property provided by DataCite schema is used to describe the relationships between the resources registered and maintained and the related resources. The schema document provides a list of control terms that can be used as values in the <relationType> property. The following is the list of allowable object type properties. When it is assumed that each property is described as ['Resource A' - Property - 'Resource B'], DataCite's ontology terms mean the following:

- IsCitedBy (A is cited by B)
- Cites (A cites B)
- IsSupplementTo (A is supplemented to B)
- IsSupplementedBy (A is supplemented by B)
- IsContinuedBy (A is continued by B)
- Continues (A continues B)
- IsNewVersionOf (A is a new version of B)
- IsPreviousVersionOf (A is a previous version of B)
- IsPartOf (A is part of B; it may be used as a property of a series element)
- HasPart (A includes B)
- IsReferencedBy (A is used as an information source of B)
- References (A uses B as its information source)
- IsDocumentedBy (B is a document describing A)

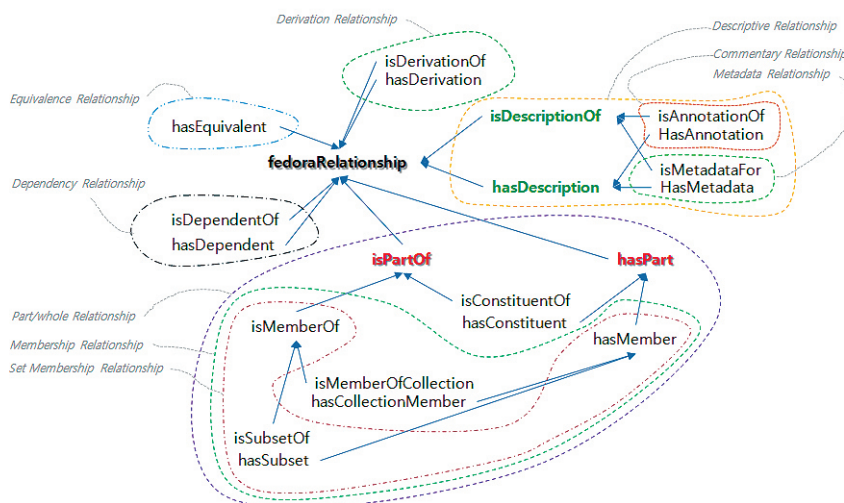


Fig. 1 Fedora Commons Object Relation Properties

⁴Starr, J., Ashton, J., Brase, J., Bracke, P., Gastl, A., Gillet, J., ... Ziedorn, F. (2011). DataCite Metadata Schema for the Publication and Citation of Research Data. Retrieved from http://schema.datacite.org/meta/kernel-2.2/doc/DataCite-MetadataKernel_v2.2.pdf

- Documents (A is a documents describing B)
- isCompiledBy (A is created through compilation by B)
- Compiles(B is created through compilation by A)
- IsVariantFormOf (A is another form of B)
- IsOriginalFormOf (A is the original form of B)

7. SCIENTIFIC DATA RESOURCE RELATION MODEL ANALYSIS

The ontology map used for Fedora Commons, DSpace, and DataCite data modeling, which is limited to the object type properties only, is illustrated in Fig. 2. There are 43 object type properties in total. Among 18 properties declared in DataCite, five properties of <isNewVersionOf>, <isReferencedBy>, <references>, <isVariantFormOf> and <isOriginalFormOf> have equivalents in the DSpace ontology.

Fig. 2 illustrates FCO, DCO, and DSO to define object type properties to describe relationship among the resources. Area (A) indicates object type properties

declared in FCO, and those written in italics mean the name of each property group. Area (B) indicates object type properties declared in DCO. Area (C) indicates object type properties declared in DSO.

In order to conduct an integrated consolidated analysis of FCO, DCO, and DSO, the relation groups and properties were studied based on five questions: 1) Do the same property names have the same meaning? 2) Are there any other property names that have the same meaning? 3) Are the relation groups of FCO applicable to DCO and DSO? 4) Is a new relation group needed? 5) Is there a need to include a property that is not used in DSO? The following is the answer to these five questions.

First, for question one, <isPartOf> and <hasPart> relations included in the Part/Whole relation group of FCO exist as the same property names with the same meanings in DCO and DSO. DCO and DSO declare <isReferencedBy> and <references> properties with the same property names and same meanings. Second, for question two, <isNewVersionOf>, <isVariant FormOf>, and <isOriginalFormOf> properties of

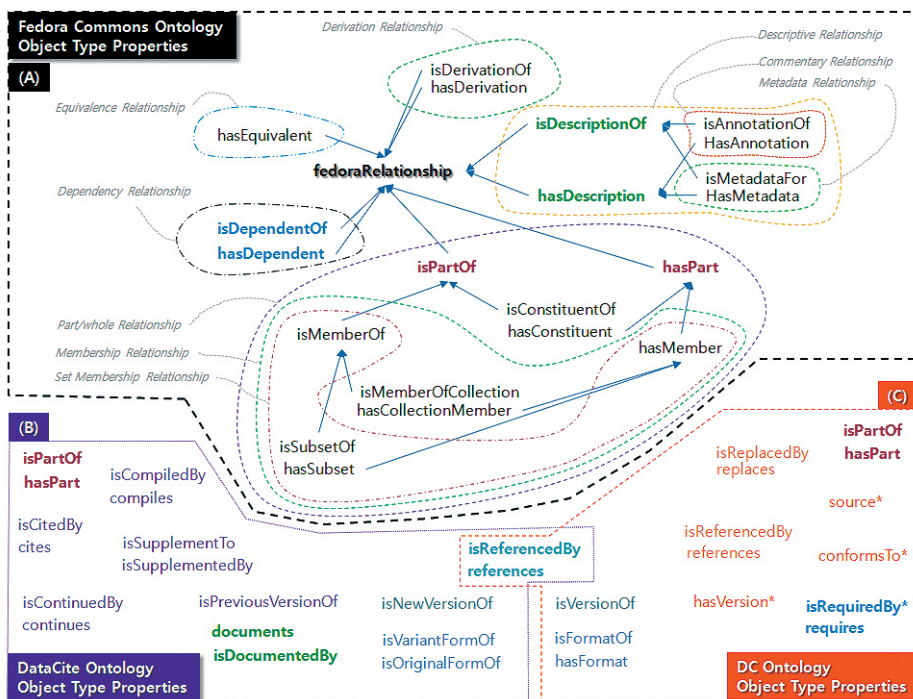


Fig. 2 The ontology map used for Fedora Commons, DSpace, and DataCite data modeling, which is limited to the object type properties only

DCO are completely equivalent to <isVersionOf>, <isFormatOf>, and <hasFormat> properties of DSO. <isDescriptionOf> and <hasDescription> properties of FCO are equivalent to the <Documents> and <isDocumentedBy> properties of DCO. Third, for question three, among object type properties declared in DCO, “isCompiledBy, compiles, isCitedBy, cites, isNewVersionOf, isPreviousVersionOf, isVariantFormOf, isOriginalFormOf, isContinuedBy, continues, isReferencedBy, references, and isSupplementTo, isSupplementedBy” relationship properties cannot be included at their current levels in the relation groups of FCO and require a separate group.

Also, among object type properties declared in DSO, “isVersionOf, hasVersion, isReplacedBy, replaces, isFormatOf, hasFormat, isReferencedBy, references, and conformsTo” relationship properties cannot be included at their current levels in the relation groups of FCO and require a separate group. “isDocumentedBy, documents, isPartOf, and hasPart” properties of DCO and “source, isRequiredBy, requires, isPartOf, hasPart”

properties of DSO, which are equivalent to those of FCO, can be included in the existing groups of the equivalent groups of FCO. Fourth, for question four, properties that do not have any equivalent property group in FCO require a separate group, but need to maintain the highest level group of the existing FCO and can be grouped through expansion of “Relation Level II.” Fifth, for the last question, “source, hasVersion, and isRequiredBy” properties of DSO are semantically interchangeable with the corresponding properties of FCO and DCO, and thus can be included. On the other hand, the <conformsTo> property does not have an equivalent property in FCO and DCO but is considered to be necessary for integration with the DSO application ontology created in various systems.

8. DISCUSSION

Table 1 shows object type properties finally derived

Table 1. Object type properties and source from FCO, DSO, and DCO

| Relation level I | Relation level II | Relation level III | Object type property | Source |
|---------------------------------|--------------------------------|------------------------------|--|-------------------|
| Derivation¹⁾ | | | isDerivationOf, hasDerivation | FCO |
| | <i>Compilation</i> | | <i>isCompiledBy, compiles</i> | DCO |
| | <i>Citation</i> | | <i>isCitedBy, cites</i> | DCO |
| | <i>Version</i> | | <i>isNewVersionOf, isPreviousVersionOf</i> | DCO |
| | <i>Replacement</i> | | <i>isReplacedBy, replaces</i> | DSO |
| Equivalence¹⁾ | | | hasEquivalent | FCO |
| | <i>Format</i> | | <i>isVariantFormOf, isOriginalFormOf</i> | DCO |
| Dependency¹⁾ | | | isDependentOf, hasDependent | FCO |
| Descriptive¹⁾ | | | IsDocumentedBy, documents | DCO |
| | <i>Commentary¹⁾</i> | | isAnnotationOf, HasAnnotation | FCO |
| | <i>Metadata¹⁾</i> | | isMetadataFor, hasMetadata | FCO |
| | <i>Continuation</i> | | <i>isContinuedBy, continues</i> | DCO |
| | <i>Reference</i> | | <i>isReferencedBy, references</i> | DCO |
| | <i>Conformity</i> | | <i>conformsTo*</i> | DSO |
| | Part/Whole¹⁾ | | | isPartOf, hasPart |
| <i>Membership¹⁾</i> | | | isMemberOfCollection, hasCollectionMember | FCO |
| | | Set Membership ¹⁾ | isSubsetOf, hasSubset | DCO |
| <i>Supplement</i> | | | <i>isSupplementTo, isSupplementedBy</i> | DCO |

to describe resource relations and their sources. Integration of the derived properties complies with DataCite property standards.

Key points of the analysis results are: 1) Compilation, Citation, Version, and Replacement relation groups are added to the Derivation relation group. As the Derivation group may include properties that include source relationship between resources, “isCompiledBy” and “compiles” properties of DCO are included in the Compilation group; “isCitedBy” and “cites” properties of DCO in the Citation group; “isNewVersionOf” and “isPreviousVersionOf” properties of DCO in the Version group; and “isReplacedBy” and “replaces” of DSO in the Replacement group. 2) The Equivalence relation group includes the Format relation group. As the Equivalence group may include properties that describe the relationship among resources with the same meaning but different formats, “isFormatOf” and “hasFormat” properties of DSO and “isVariantFormOf” and “isOriginalFormOf” properties of DCO are included in the Format group. 3) To the Descriptive relation group, Continuation, Reference, and Conformity relation groups are added.

The properties included in the Descriptive group describe additional information about resources. For example, they may include metadata that describes raw data rather than relationships among raw data, and properties that describe annotations. Therefore, “isContinuedBy” and “continues” properties of DCO are included in the Continuation group, and “isReferencedBy” and “references” properties declared both in DSO and DCO are included in the Reference group. <conformsTo> of DSO, the property describing the relation with the standards or guidelines that the resources conform to, is included in the Conformity group. 4) To the Part/Whole relation group, the Supplement relation group is added. As “isSupplementTo” and “isSupplementedBy” properties of DCO can be included in the Part/Whole group, the Supplement relation group is included as a sub-group of the Part/Whole relation group and is also included in the Supplement relation group.

In this paper, the suggested object type properties are composed of five terms in Relation level I, twelve terms in relation level II, one term in relation level III, and thirty-four terms in object type properties. Not all

object type properties in FCO, DSO, and DCO are omitted because the object type properties are examined by the meaning of the terms and the representative terms are deducted.

KISTI is planning to disseminate the P-CUBE to the KOPRI (Korea Polar Research Institute). The data which are created and assimilated in this discipline may have specific object type properties. As mentioned already in the RESEARCH METHODOLOGY section, the property analysis was conducted limited to the object type properties that describe relationships among the resources. The data type properties for ontology are not analyzed because each disciplinary area has their own properties to describe their data in their fields. So additional studies for domain specific data type properties and object type properties are needed.

9. CONCLUSIONS

Along with the development of hardware and software and evolution of the network environment, various forms of data have been produced in large quantities. As people have become more aware that the research data produced with the nation's investments in research and development are national assets, there have been active movements in developed countries to collect, manage, conserve, and service raw data to validate the literature-based research results. In this regard, research efforts about metadata to describe resources have been constantly carried out. In this study, the object type properties required to describe the relationship among research data resources were derived.

In this study, the object type properties declared in FCO, DSO, and DCO were analyzed, and then integrated. The systems which use the Fedora Commons or DSpace can be expanded by using this ontology to describe relations among the managed data. The properties of Fedora Commons and DSpace, which are open source software platforms for resource management, and the schema properties published in the DataCite community were analyzed. Based on the relation group names of Fedora Commons, nine new relation group names including Format were derived.

Through integration of the properties, 38 object type properties including <isDerivationOf> were derived.

The groups and the object type properties derived in this study can be used to describe the relationships among resources for resource management in various fields, without being limited to certain domains.

REFERENCES

- Goranova, M., Shishedjiev, B., & Georgieva, J. (2011). Research on Building Scientific Data Ontology. *2011 Developments in E-systems Engineering* (pp. 541-546). Dubai: IEEE Press.
- Soldatova, L. N., & King, R. D. (2006). An ontology of scientific experiments. *Journal of the Royal Society Interface* 3, 795-803.
- Shotton, D. (2011). DataCite2RDF—Mapping Data Cite Metadata Scheme Terms to ontologies. Retrieved from <http://opencitations.wordpress.com/2011/06/30/datacite2rdf-mapping-datacite-metadata-scheme-terms-to-ontologies-2/>
- Washington, N., & Lewis, S. (2008). *Ontologies: Scientific Data Sharing Made Easy*. Retrieved from <http://www.nature.com/scitable/topicpage/ontologies-scientific-data-sharing-made-easy-77972>.

APPENDIX

The table below shows object type properties to describe resource relations and the group of each property. For expression of the relation name, the relationships noted with the superscript '1)' comply with the relation group expression of FCO. The properties noted with '*' mean the properties declared in the Dublin Core-Library Application Profile, though not declared in DSO. Each relation level has a hierarchical structure. Property names defined in FCO, DSO, and DCO have no hierarchical structure. Relations shown in the same column indicate that these properties have the same meanings. Newly included property names and the corresponding object type properties are noted in italics.

Crosswalk for FCO (Fedora Commons Ontology), DSO (Dspace Ontology), DCO (DataCite Ontology). Relations I, II, III show the relation depth.

| Relation level I | Relation level II | Relation level III | FCO | DSO | DCO |
|---------------------------------|--------------------|--------------------|---------------------------------|--|---|
| Derivation¹⁾ | | | isDerivationOf hasDerivation | source* | |
| | <i>Compilation</i> | | | | <i>isCompiledBy</i> <i>Compiles</i> |
| | <i>Citation</i> | | | | <i>isCitedBy</i> <i>cites</i> |
| | <i>Version</i> | | | <i>isVersionOf</i> <i>hasVersion*</i> | <i>isNewVersionOf</i> <i>isPreviousVersionOf</i> |
| | <i>Replacement</i> | | | <i>isRequiredBy</i> <i>replaces</i> | |
| Equivalence¹⁾ | | | hasEquivalent | | |
| | <i>Format</i> | | | <i>isFormatOf</i> <i>hasFormat</i> | <i>isVariantFormOf</i> <i>isOriginalFormOf</i> |
| Dependency¹⁾ | | | isDependentOf | isRequiredBy* | |

