1st Conference on Research Data Infrastructure

Enabling RDM

https://doi.org/10.52825/CoRDI.v1i.260

Published: 07 Sept. 2023

Toward the development of NII RDC application profile using ontology technology

Yasuyuki Minamiyama[1][https://orcid.org/0000-0002-7280-3342], Masaharu Hayashi[1][https://orcid.org/0000-0002-1451-9300], Ikki Fujiwara [1][https://orcid.org/0000-0001-6841-5243], Jun-ichi Onami [1][https://orcid.org/0000-0003-0790-8313], Shigetoshi Yokoyama [1][https://orcid.org/0000-0003-0060-8166], Yusuke Komiyama [1][https://orcid.org/0000-0001-6468-3718], and Kazutsuna Yamaji[1][https://orcid.org/0000-0001-6108-9385]

[1] National Institute of Informatics, Japan

Abstract. In recent years, there has been international progress in developing platforms that support the reproducibility and reusability of research data. Typical platforms adopt a service architecture integrating multiple information systems to cover the entire research data lifecycle. In realizing this architecture, specifications for inheriting processes and results executed on different information systems play an essential role. This study introduces our practices for application profile development using ontology technology in the NII Research Data Cloud.

Keywords: Research Data Platform, Application Profile, Interoperability, Ontology

1. Introduction

In recent years, there has been international progress in developing platforms that support the reproducibility and reusability of research data. For example, EUDAT adopts a service architecture integrating multiple information systems to cover the entire research data lifecycle [1]. Even in Japan, the NII Research Data Cloud (NII RDC) has been under development since 2017, led by the National Institute of Informatics [2].

In realizing this architecture, ensuring data interoperability is becoming the next challenge. If the meaning and structure assigned to research data generated by one information system cannot be adequately inherited, a different information system cannot interpret the data. This failure causes a severe loss of reproducibility and reusability of the research.

To address this issue, developing an application profile is essential in handling the meaning and structure of data between different information systems [3]. In RDM platform development, research data interoperability could be maximized by developing application profiles. However, there is no established procedure for constructing application profiles. This study introduces the process for developing an application profile using ontology theory and technology through our NII RDC development experience.

2. Approach

This chapter discusses the application profile development in line with ontology theory. We adopted the seven steps proposed by Noy & McGuiness as a development process [4]. We introduce “Step 3: Enumerate important terms in the ontology” and “Step 4: Define the classes and the class hierarchy” as distinctive steps in this paper.

2.1 Step 3: Enumerate important terms in the ontology

This step extracts the key terms to be addressed in the application profile. As a premise, we use “user story” method for determining specific functional requirements for NII RDC. A user story is a description of a function that is valuable to the user of the system or software, as expressed in the following manner:

"As [a user persona], I want [to perform this action] so that [I can accomplish this goal]"

These stories include essential information for understanding the meaning and structure assigned to the data generated by an information system. We extracted necessary terms from 138 NII RDC user stories by splitting them into subject/verb/object.

2.2 Define the classes and the class hierarchy

This step defines the classes and hierarchical relationships for this application profile. In designing the classes and hierarchical relationships, we extended the Activity Streams 2.0 framework. Activity Streams 2.0 enables the research data lifecycle to be viewed as a series of “RDM-related activities." Table 1 shows the lists of the defined classes.

Table 1. List of the defined classes.

Types	**Class**	**Definition**
Actor Types	Person	https://www.w3.org/ns/activitystreams#Person
	Institution	https://www.w3.org/ns/activitystreams#Organization
	Funding Agency	https://www.w3.org/ns/activitystreams#Organization
	Application	https://www.w3.org/ns/activitystreams#Application
	Service	https://www.w3.org/ns/activitystreams#Service
Activity Types	Activity	https://www.w3.org/ns/activitystreams#Activity ; **https://purl.org/rdm/ontology**
Object and Link Types	Resource	https://www.w3.org/TR/activitystreams-vocabulary/#dfn-object
	Repository	https://www.w3.org/TR/activitystreams-vocabulary/#Place
	Project	https://www.w3.org/ns/activitystreams#context
	DataManagementPlan	https://www.w3.org/ns/activitystreams#Document
	Collection	https://www.w3.org/ns/activitystreams#Collection
	Event	https://www.w3.org/ns/activitystreams#Event
	Access Rights Information	http://purl.org/dc/terms/RightsStatement
	Document	https://www.w3.org/ns/activitystreams#Document

We assigned "Actor Types" as the subject, "Activity Types" as the verb, and "Object and Link Types" as the target. Also, we defined 72 verbs extracted in Section 2.1 as an extension of the "Activity" class. The URIs and definitions of each vocabulary are available at https://purl.org/rdm/ontology. Note that "access rights" was found to be a term not covered by Activity Streams 2.0, so we adopted "Rights Statement" defined by DCMI Metadata Terms.

3. Implementation

We described the NII RDC user story in JSON format based on the application profile developed in Chapter 2. Figure 1 shows an example.

```
EXAMPLE XX

{
  "@context": [
    "https://www.w3.org/ns/activitystreams",
    "https://purl.org/rdm/ontology"
  ],
  "summary": "As a Principal Investigator, I want to create an appropriate Data Management Plan so that I can comply the requirement by Funding Agency.",
  "type": "create",
  "actor": {
    "id": "https://orcid.org/0000-0002-7280-3342",
    "type": "Person",
    "name": "Yasuyuki Minamiyama",
    "role": "ProjectLeader"
  },
  "object": {
    "id": "https://doi.org/10.20736/12345678",
    "type": "DataManagementPlan",
    "name": "NII RDC project",
    "published": "2023-03-05T00:00:00"
  },
  "instrument": {
      "type": "Service",
      "name": "GakuNin RDM"
    },
}
```

Figure 1. Example description based on NII RDC application profile.

The structure of "type," "actor," "object," and "instrument" is in line with Activity Streams 2.0. The original user stories are described as "summary" so that the correspondence can be checked in case of later modification.

4. Conclusion

Implementation based on this application profile will facilitate communication between NII RDC systems and record the semantics contained in their activities. This approach can potentially improve data interoperability and be used as primary data for constructing a knowledge graph that represents RDM. We plan to explore the possibility of utilization through future analysis.

Underlying and related material

The application profile and the related document will be available at the following URL: https://purl.org/rdm/ontology.

Author contributions

Y.M. substantially contributed to the study conceptualization and the design of methodology. M.H., I.F., J.O., S.Y., and Y.K. contributed to the validation of the results. K.Y. supervised and directed the project. All authors discussed the results and contributed to the final manuscript.

Competing interests

The authors declare that they have no competing interests.

References

1. S. de Witt, D. Lecarpentier, M. van de Sanden, and J. Reetz. “EUDAT - A Pan-European Perspective on Data Management,” in 2017 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), pp. 1–5, 2017.
2. Research Center for Open Science and Data Platform (RCOS), National Institute of Informatics. “Overview of the NII Research Data Cloud,” https://rcos.nii.ac.jp/en/service/, (accessed 2023-4-6).
3. Rachel Heery, Manjula Patel. “Application profiles: mixing and matching metadata schemas,” Ariadne Issue 25, (2000), (accessed 2023-4-6).
4. Noy, N.F., & McGuinness, D.L. “Ontology Development 101: A Guide to Creating Your First Ontology,” 2001, https://protege.stanford.edu/publications/ontology_development/ontology101.pdf, (accessed 2023-4-6).