On FAIR Data Principles of Institutional Data and Information of Universities

MORI Masao (1), OISHI Tetsuya (2), TAKATA Eiichi (3), OGASHIWA Kahori (4), SHIRATORI Naruhiko (5), TAJIRI Shintaro (6)

Tokyo Institute of Technology, Japan (1)
Tokyo Institute of Technology, Japan (2)
Kobe University, Japan (3)
Teikyo University, Japan (4)
Kaetsu University, Japan (5)
Yokohama College of Commerce, Japan (6)

Abstract

In scientific research, it costs a lot to reproduce experiments succeeded in the past. For efficiency of research, scientific researchers come to reuse research data of the past research. To do so, it is necessary not only to archive research data, but research data should be arranged confirming to FAIR data principles. FAIR is acronym of "findability", "accessibility", "interoperability" and "re-usability". To realize FAIR data principle, fulfillment of metadata and persistent identifier on data is significant. On the other hands, institutional data on education and institutional management are coming to be recognized significant for university management to secure fundamental information of universities. In that context the FAIR data principle is effective to facilitate collecting institutional data as well. However, generally speaking, as the technical situations of institutional information are different from university to university, it is difficult to acquire a general resolution. We study the problem to make institutional data and information to be abstract using ontology engineering, and consider the design of institutional data and information that meets FAIR data principles.

Keywords: *FAIR data principle, open science, institutional research, ontology;*

1. Introduction

FAIR Data Principle (Fig1) has been proposed by FORCE11, the community for research academics, librarian and research funders, aiming at good data management for scientific research data [1]. From 2011, FORCE11 started to discuss and output the guide line in 2016. EU adopted FAIR Data Principle for Open Data Policy Pilot.

"FAIR" is an acronym for four attributes of scientific research data; to be Findable, to be Accessible, to be Interoperable and to be Re-usable. In order to realize the principles, research data need to be indexed by PID (persistent identifier), to be enriched with meta data, to be deployed on database system, and to be on networks.

FAIR Data Principles

TO BE FINDABLE:

F1. (meta)data are assigned a globally unique and eternally persistent identifier.
F2. data are described with rich metadata.
F3. (meta)data are registered or indexed in a searchable resource.
F4. metadata specify the data identifier.

TO BE ACCESSIBLE:

A1 (meta)data are retrievable by their identifier using a standardized communications protocol.
A1.1 the protocol is open, free, and universally implementable.
A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
A2 metadata are accessible, even when the data are no longer available.

TO BE INTEROPERABLE:

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
I2. (meta)data use vocabularies that follow FAIR principles.
I3. (meta)data include qualified references to other (meta)data.

TO BE RE-USABLE:

R1. meta(data) have a plurality of accurate and relevant attributes.
R1.1. (meta)data are released with a clear and accessible data usage license.
R1.2. (meta)data are associated with their provenance.
R1.3. (meta)data meet domain-relevant community standards.

Fig. 1 FAIR Data Principles

Institutional Research is aiming at supporting decision making and planning for universities and research institutions by analyzing data analysis and supplying strategic information. The data and information are required to be comprehensive, exhausted, period-continuous and indexed as possible. As FAIR Data Principles meet these requirements, FAIRness is necessary principles on institutional data and information for institutional research.

This study gives a method to describe metadata of institutional information and data using ontology in information science.

2. Ontology in Information Science

A “Domain knowledge” is a system of knowledge containing, for instance, within computer software or business procedure. Ontology in information science is a technique to describe formal representation of concepts and relationships between those concepts in domain knowledge. For instance, Common European Research Information Format (CERIF) has defined an ontology in the form of diagram like (Fig.2), but normally ontology is described in XML format.

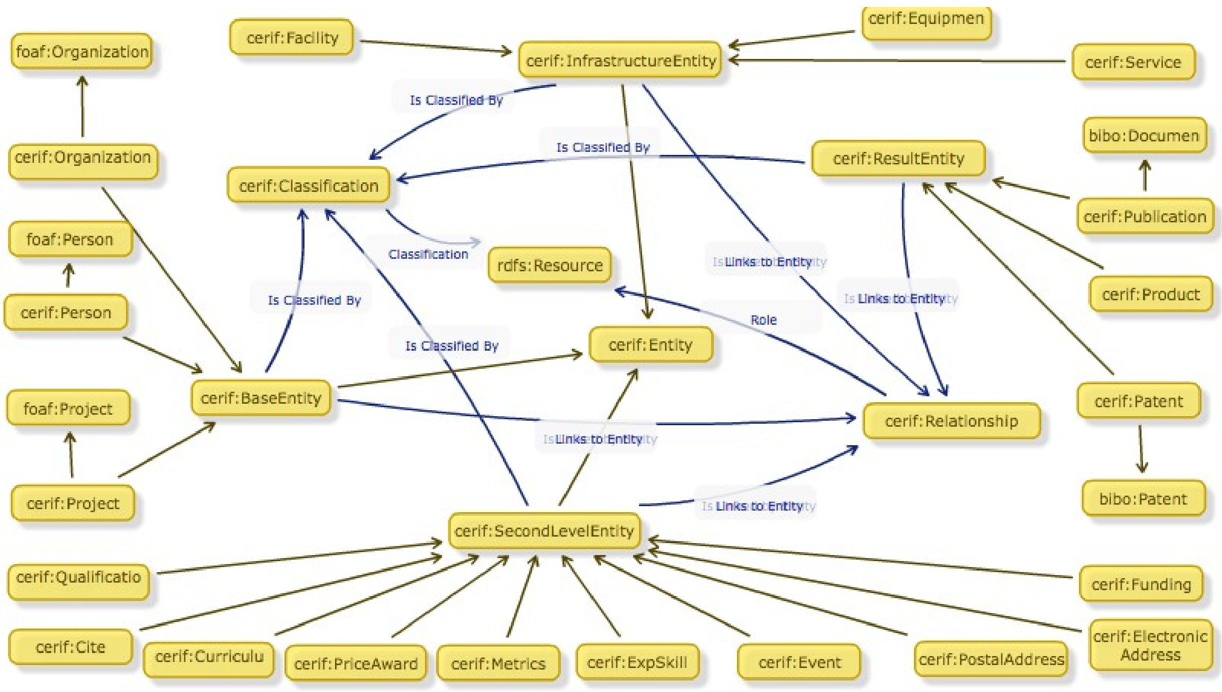

Fig. 2 Ontology of CERIF 1.3 (Quoted from [3])

In ontology engineering, concepts in domain knowledge are called ‘Class’. Class has rank, which is called ‘Relation’. ‘Is-a relation’ expresses relation between upper class and lower class. In some case, a set of classes is described as a class, then we have ‘part-of relation’ which means relationship between set-class of classes and its element-class. Attribute of class is expressed ‘attribute-of relation’.

If the relationship between the classes is kept to the minimum necessary, the designer may arbitrarily add. Generally, it is said that there is no unique way to create an ontology for any knowledge system, and from various viewpoints utilizing the ontology to be constructed, it is only necessary to make it appropriate system according to necessity [2]. Therefore, the ontology of university information shown here is one possibility, not a suggestion as the only solution.

For example, in the ontology of CERIF in Fig. 2, the class cerif:Publication (publication) is a upper class of cerif:ResultEntity (is-a relation), and arrows are drawn from the lower class to the upper class. Furthermore, cerif:ResultEntity is one of the items in cerif:Classification (classification of research information) (is-class fied-by relationship) (see Fig. 3). The last is-classified-by relationship is an additionally introduced class relation. This is an example in which relationship can be flexibly introduced in ontology engineering. This ontology represents the following description:

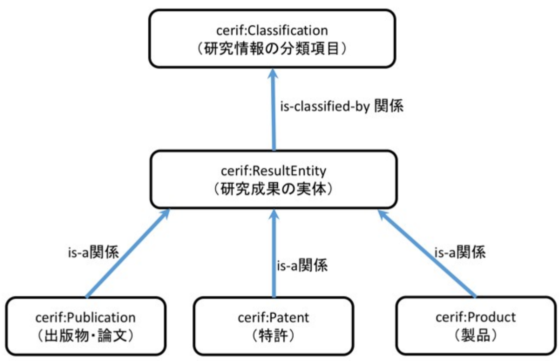

Fig. 3 Part of ontology of CERIF

Research results consist of papers, publications, products and patents. The research outcome is one item of research information, and information on basic object (BaseEntity) and infrastructure (InfrastructureEntity) are also included as items of research information.

We can see that ontology engineering is used to expresses concepts and relations efficiently in the domain knowledge of CERIF.

3. Construction of The Ontology for Institutional Data : example

In this section, as an example of an ontology, we prototype an ontology of information on education and its outcome. The following six items are listed as possible classes.

1. Student
2. Faculty
3. Department
4. Lesson
5. Curriculum
6. Grade

Lesson and Curriculum are clearly an is-a relationship. Looking at the ontology of CERIF, engagement of Organization and Person are comprehended by introducing an upper class called BaseEntity.

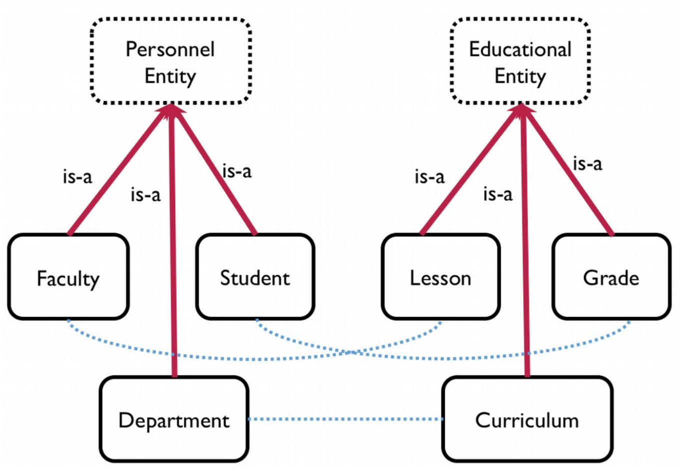

Fig. 4 Example of ontology on institutional information

Following this, it is conceivable that for Student, Faculty and Department we can construct unification-class by introducing upper classes like PersonnelEntity. In summary, we can propose the ontology as shown in Fig. 4. Solid arrows indicate. The relationship between the classes mentioned above is shown, and the broken line shows what is undefined due to the relationship of the class which needs to be considered.

4. Conclusion

In this paper, in order to apply FAIR Data Principles to institutional data and information, it is necessary to describe those metadata. The method of ontology engineering is helpful to understand the structure of metadata. As an advanced example, we introduced CERIF ontology. Based on the ontology structure used there, we examined the method to be carried out in this research in the future.

References

[1] FORCE11, “The fair data principles”, https://www.force11.org/group/fairgroup/fairprinciples, 2016
[2] Natalya F. Noy and Deborah L. McGuinness, “Ontology Development 101: A Guide to Creating Your First Ontology”, Stanford Knowledge Systems Laboratory Technical report KSL-01-05, 2001
[3] Iván Ruiz Rube, Keith G Jeffery, Jan Dvorak, Brigitte Joerg, Geert van Grootel, Miguel Ángel Sicilia, “CERIF Ontology 1.3”, 2013.