Metadata-driven interdisciplinary research projects using RIKEN MetaDatabase

Norio Kobayashi[1,2,3], Satoshi Kume[3], and Hiroshi Masuya[2,1]

[1] Advanced Center for Computing and Communication (ACCC), RIKEN,
2-1 Hirosawa, Wako, Saitama, 351-0198 Japan
norio.kobayashi@riken.jp
[2] BioResource Center (BRC), RIKEN,
3-1-1, Koyadai,Tsukuba, Ibaraki, 305-0074 Japan
hmasuya@brc.riken.jp
[3] RIKEN CLST-JEOL Collaboration Center, RIKEN,
6-7-3 Minatojima-minamimachi, Chuo-ku, Kobe, Hyogo, 650-0047 Japan
satoshi.kume@riken.jp

Abstract. The life sciences are data driven. However, as life science datasets are highly heterogeneous, the comprehensible data publication, published data discovery and data integration are urgent problems for researchers. Our database publication platform called RIKEN MetaDatabase was developed to solve the problems in the metadata approach based on the resource description framework (RDF). As of September 2017, the platform has 26 ontologies and 135 databases in various life science fields, and provides a simple graphical user interface that has a tabular view. Recently, by using the platform, we have developed an electron microscopy image database with a specialised image viewer and started several engineering research projects including health care, sustainable humanoshere, and material and measurement categories. This paper presents these interdisciplinary applications of the RIKEN MetaDatabase.

Keywords: Semantic Web, database platform, database integration, life sciences, metadata-driven research project

1 Introduction

In the life sciences, apparatus measuring various biological phenomena are rapidly being developed, and huge and heterogeneous datasets are being created. Although the primary idea of life sciences is to understand the phenomenon of life by multilateral data analysis, researchers face great difficulties in integrative analysis and in discovering and publishing such heterogeneous datasets. Inside RIKEN, the largest Japanese comprehensive science institute, which has large-scale research centres and many small-scale laboratories, is facing the same problem. To address this issue, we developed a database platform called RIKEN MetaDatabase (`http://metadb.riken.jp`) based on the resource description

framework (RDF), which enables researchers to integrate and publish their data on the Web. The platform collects individual databases created by researchers or research projects, and integrates them by sharing ontologies and RDF resources.The RIKEN MetaDatabase was launched in April 2015, and 135 databases, including RIKENs original 61 databases and 26 public ontologies have been integrated and published on the platform. In the following section, we introduce our application and research projects using this platform.

2 Application: a microscopy imaging viewer

The RIKEN Microstructural Imaging Metadatabase (`http://clst.multimodal.riken.jp/CLST_ManageData/RikenImageDB/`) implements a specialised graphical user interface and image viewer, which can be accessed through a web browser. It acts as a SPARQL client for the RIKEN MetaDatabase to obtain metadata as image annotation data. By using such metadata, the viewer generates a list of image data for each organ and species, loads a Deep Zoom Image (DZI), and displays the DZI with associated metadata, including the imaging conditions and biosample. We developed a workflow to generate metadata immediately after the imaging process by collaboration among wet and dry researchers. We have published scanning electron microscope (SEM) images of rat livers and kidneys as well as images of human blood cells.

3 Metadata-driven engineering research projects

We have just started new research projects, which are listed below, and which aim to promote technological innovation that will contribute to humanity and society by developing an engineering oriented researcher network, called Engineering Network. (1) Health care: We have been developing high-speed accurate cytology for leukaemia diagnosis using microstructure imaging. This system will automatically diagnose the type of leukaemia in detail by analysing a SEM image showing thousands of blood cells with metadata using artificial intelligence such as deep learning. (2) Sustainable humanoshere: Toward a comprehensive solution for Sustainable Development Goals, this project introduces a novel methodology that reconnects and controls the environment and human society resource cycles. More concretely, it introduces the high-precision computer simulation of biomass utilisation technology, including introduction of profitable insects over biomass, such as fallen leaves and branches, which have been measured in our institute and precisely annotated by metadata. (3) Material and measurement: In the development of a data driven manufacturing system, the goal is to develop a prototype processing system for novel materials such as bio-plastic and novel processing methods, which quickly learn suitable and adaptable methods with fewer trials by implementing a metadata database of material properties, and processing characteristics and computer simulations using the metadata.

These research projects are original in the sense that they are metadata-driven engineering projects that utilise the RIKEN MetaDatabase from the beginning. Domain researchers and informaticians collaborate and generate highly accurate metadata from the beginning of these interdisciplinary projects.