Title	
Author(s)	,
Citation	
Issue Date	2023-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/18297
Rights	
Description	Supervisor: , , (

JAIST

JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

Abstract

In recent years, as exemplified by the government-led data utilization in Society 5.0, it is expected that the sophistication and efficiency of production activities and the convenience of services will be improved by linking and utilizing data dispersed and held by the national government, local governments, private companies, and other entities.

However, there have been issues in terms of interoperability, such as the lack of uniformity in the notation of information in data catalogs in the linkage of data from various fields in industry, government, and academia. In response to this situation, the Data Society Promotion Council is promoting guidelines for data catalog creation in order to utilize sensing data and device data. This will enable standardization of description methods when explaining metadata, and will make it possible to search for data using metadata and link data between fields in the future.

In data utilization, there is an increasing trend toward the use of not only primary data, which is raw data directly collected from IoT devices, etc., but also secondary data generated by processing multiple pieces of primary data, etc. The number of combinations of original data and processed data is increasing. As the number of types of secondary data increases with the number of combinations of original data and processing, it may become difficult to manually assign metadata to the secondary data. In addition, how to assign metadata is becoming a problem.

In addition, in the case of complex structures that have been processed multiple times, it is impossible to obtain information on the original data and processing from the last generated data, and it is difficult to know how the secondary data was generated.

In this study, we proposed a system to solve these problems. As a proposed system, for the problem of secondary data that increases with the number of combinations, the processor creates a program for metadata generation when the primary data is processed and secondary data is generated. This will allow metadata to be assigned at a stage before the number of types increases, thus solving the problem of secondary data that continues to increase with the number of combinations. To address the issue of the features of the generated secondary data, metadata of the primary data used in the processing process can be added to the metadata of the secondary data as a history to explain the processing route, etc. of the generated secondary data. We believe that this will allow secondary data to be registered in data distribution.

By utilizing the proposed system, we believe that users will be better served by being able to provide secondary data. We also believe that the data distribution market can be expected to be activated as these data become even more available.