

Development of an Information Research Platform for Data-Driven Agriculture

Takahiro Kawamura, National Agriculture and Food Research Organization, Japan*

Tetsuo Katsuragi, National Agriculture and Food Research Organization, Japan

Akio Kobayashi, National Agriculture and Food Research Organization, Japan

Motoko Inatomi, National Agriculture and Food Research Organization, Japan

Masataka Oshiro, National Agriculture and Food Research Organization, Japan

Hisashi Eguchi, National Agriculture and Food Research Organization, Japan

ABSTRACT

Comprehensive research data are acknowledged as a necessity for research acceleration. Research institutes and universities are engaged in developing research data management systems. The National Agriculture and Food Research Organization of Japan (NARO) developed NARO-linked databases (Narolin DBs) in addition to a supercomputer. In the Narolin DB various research data on agriculture are cataloged using common metadata. The relationship between complicated data in natural science is described in RDF, property graph, or RDB format to facilitate the application of statistical analysis and machine learning. The system is unique in that it is connected to a data catalog, a private cloud database, a supercomputer for data analysis, and a data/service portal for business applications, such as a data pipeline. Through the development of agricultural information research platforms, NARO will accelerate data-driven agricultural research at various stages in the agricultural supply chain, ranging from genome analysis to plant breeding, cultivation, food processing, and food distribution.

KEYWORDS

Data Catalog, Research Data Management, Smart Agriculture

INTRODUCTION

In recent years, information and communication technology (ICT) use has increased, even in the field of agriculture, and research data have been digitized in large volumes. By appropriately collecting and managing these research data, sharing them to the extent necessary, and applying data science and AI technology, such as statistical analysis and machine learning, higher efficiency and profitability of agriculture and the creation of interdisciplinary agricultural research beyond the conventional field and region is expected. These processes are called “smart agriculture” or “data-driven agriculture,” which are also required to strategically utilize research data that have not been previously shared by opening the data to external organizations, especially in the area of international cooperation (open data), sharing it among consortium parties, such as national projects (disclosed data), or concealing

DOI: 10.4018/IJAEIS.302908

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

it for competitive utilization (closed data). Smart agriculture and data-driven agriculture have been around for a while; however, these processes are still not widely used, and the use of data is insufficient. One of the reasons for this is that many people, such as researchers, agricultural companies, farmers, do not know what data is available, what it is for, and how it can be used. So, the research issue in this paper is: How can we make better use of data in the agricultural sector? In other words, as a national research institute, what kind of comprehensive mechanism should we prepare to promote data utilization?

The National Agriculture and Food Research Organization (NARO), the largest agricultural research body that is composed of several institutes, was established in FY 2016 through the integration of multiple institutes under the Ministry of Agriculture, Forestry and Fisheries (MAFF) in Japan. NARO found it necessary to establish a unified research data strategy and infrastructure to promote data-driven agricultural research and development while also coordinating research data across diverse research fields, such as genomic information on animals, plants, and microorganisms, breeding information, cultivation and growth management information, including viruses and pest control, food processing and distribution information, robotic tractor design, and environmental information, such as climate change and soil information. According to a preliminary survey conducted in 2019, approximately 60% of the research data were stored in the researcher's PCs and Hard Disk Drives (HDDs) causing data sharing inside and outside NARO to be delayed.

Therefore, NARO developed an agricultural information research platform, including NARO linked databases on agricultural-related data (Narolin DB), a supercomputer that utilized the database for AI operations such as machine learning and a data distribution system called WAGRI. Especially, in response to the above research issue, a platform was provided as a pipeline to collect data, add unified metadata to make it easy to discover in the primary DB, prepare the data format as pre-processing for analysis in the secondary DB, and make it easy for non-ICT experts in the agricultural sector to utilize the data by data analysis environment provided by AI supercomputer, and by WAGRI to provide data via API to private companies that serve farmers as end-users. In addition to the development of such computational systems, an integrated mechanism was created that included institutional design for human systems, such as formulation of terms of use for research data and AI education. Thus, the contribution in this paper is to share the knowledge and experience acquired in the design and implementation of the comprehensive mechanism to promote data utilization in the agricultural sector.

In the rest of this paper, the systems and services related to the research data are outlined in the second section. The structure of NARO's information research platform and the common metadata of NARO are described in the third section. After the introduction of a supercomputer for AI operations in the fourth section, and WAGRI for data distribution for business in the fifth section, the authors discuss the comparison with existing systems and case studies on their platform in the sixth section, and a summary, future issues, and institutional measures for dissemination of the data-driven approach are explored in the final section.

RELATED WORK

A research data catalog or data repository, in a broad sense, is an information platform for electronic data storage and sharing. In particular, it refers to a system established by national research institutes, universities, and related organizations, which store and manage research data obtained with public funds in a form that can be reused, and additionally, the system provides data utilization services inside and outside the institutions (CabinetOffice, 2019). Apart from those individually provided as services by national institutes and universities, systems and software that can be used by third parties include CKAN, DKAN, Dataverse, Open Science Framework, figshare, data.world, Mendeley Data, Dryad, Open ICPSR, Zenodo, and others. In a July 2017 article (Dataverse, 2017, July 25), there was a comparison of the functionality of nine research data catalogs. In the following section, four systems, especially related to Narolin DB, and some other services, are introduced.

17 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/development-of-an-information-research-platform-for-data-driven-agriculture/302908

Related Content

Use of Geo-Information in Environmental Policy: Limitations, Advantages and Challenges

Wies Vullings, Jandirk Bulensand Dennis Walvoort (2014). *International Journal of Agricultural and Environmental Information Systems* (pp. 38-58).

www.irma-international.org/article/use-of-geo-information-in-environmental-policy/111216

Provincial Linkage Characteristics of Hog Price in China Based on Linkage Social Network Analysis Method

Jiyun Bai, Muyan Liu, Li Maand Jun Meng (2020). *International Journal of Agricultural and Environmental Information Systems* (pp. 61-74).

www.irma-international.org/article/provincial-linkage-characteristics-of-hog-price-in-china-based-on-linkage-social-network-analysis-method/256991

Implementation of Improved Control Strategy of DC-AC Converter using Delta-Sigma Modulator

Mohamayee Mohapatraand Chitti Babu Baladhandautham (2016). *Smart Grid as a Solution for Renewable and Efficient Energy* (pp. 249-272).

www.irma-international.org/chapter/implementation-of-improved-control-strategy-of-dc-ac-converter-using-delta-sigma-modulator/150324

Optimization Studies on Biosorption of Ni(ii) and Cd(ii) from Wastewater in a Packed Bed Bioreactor

K. Narasimhuluand Y. Pydi Setty (2015). *Handbook of Research on Uncovering New Methods for Ecosystem Management through Bioremediation* (pp. 367-398).

www.irma-international.org/chapter/optimization-studies-on-biosorption-of-niii-and-cdii-from-wastewater-in-a-packed-bed-bioreactor/135103

A Strategy for Estimating Nutrient Concentrations Using Remote Sensing Datasets and Hydrological Modeling

Vladimir J. Alarconand William H. McAnally (2012). *International Journal of Agricultural and Environmental Information Systems* (pp. 1-13).

www.irma-international.org/article/strategy-estimating-nutrient-concentrations-using/62062