REVIEW

Common Agriculture Vocabulary for Enhancing Semantic-level Interoperability in Japan

Akane TAKEZAKI[1*], Sungmin JOO[2], Hideaki TAKEDA[2] and Tomokazu YOSHIDA[1]

[1] National Agriculture and Food Research Organization, Tsukuba, Japan
[2] National Institute of Informatics, Chiyoda, Tokyo, Japan

Abstract

We introduce the features of Common Agriculture Vocabulary (CAVOC) and CAVOC-based services, review other resources for standards in the domains of agriculture and food, and briefly discuss the future direction of agricultural semantics standards in Japan based on reviews. CAVOC has been developed as a core vocabulary for enhancing semantics-level interoperability. It specifies concepts that are frequently used to annotate agricultural data on agricultural activities and crops in Japan. The CAVOC website provides uniform resource identifiers by which the concepts in CAVOC are given. The data in CAVOC can be downloaded from the website. The extension of CAVOC coverage is important to unambiguously define objects and their relationships in the domains of agriculture and food. International collaborative efforts to create standards in agricultural semantics are underway through mapping among semantic resources within international projects and by holding meetings among stakeholders. We need to join the international community in creating agricultural semantics standards so as to extend CAVOC coverage through the utilization of foreign resources.

Introduction

Large amounts of data in the domains of agriculture and food are now being collected from such diverse sources as sensors, satellites, and drones, and used not only to support decision-making to maximize farm profitability and optimize farm management, but also to identify the origin of products and thus secure product safety. Analysis of data from diverse sources can also provide contextual and situational awareness in response to real-time events such as weather or disease alerts, and can thus help in making agile decisions in response to sudden changes in operational conditions (Wolfert et al. 2014). One of the bottlenecks hindering the promotion of data-oriented farming is a lack of interoperability among agricultural information and communications technology (ICT) systems. Inconsistencies in terminology and meanings result in poor semantic interoperability (Zeng 2019). For example, "1.5 h" may be recorded as the time required for "basal dressing" in one system, whereas "2.5 h" may be required for "basal application" in another system. When an attempt is made to retrieve information across these different systems, "2.5 h" for "basal application" will not be included in the results of the search query "basal dressing" and consequently, the two data cannot be compared. The Cabinet Office of Japan has published guidelines on "Standard Terms of Use in Agricultural IT Services" (https://www.kantei. go.jp/jp/singi/it2/senmon_bunka/nougyou.html, accessed 30 July 2019), and promoted the use of these standard terms to secure semantic interoperability. However, it has been difficult to unify the terms used in agricultural ICT systems, because most agricultural terms are specific to the rural area, crop, or cultivation

method used. We therefore developed the Common Agriculture Vocabulary (CAVOC) to achieve semantic interoperability. CAVOC explicitly defines the relationships of standard terms in the Cabinet Office's guidelines with other terms, converting them into the standard terms. In the example given above, if the information retrieval system refers to CAVOC and finds that "basal dressing" and "basal application" have the same meaning, then "2.5 h" for "basal application" will be included in the results of a search using the query term "basal dressing."

Here, we introduce the features of CAVOC and CAVOC-based services, review other resources for standards, information systems with semantic technologies, and international collaborative efforts for standards in the domains of agriculture and food, and then briefly discuss the future direction of our work based on this review.

Features of CAVOC and CAVOC-based services

1. CAVOC features

CAVOC specifies concepts that are frequently used to annotate agricultural data on agricultural activities and crops. These concepts are included in the "Standard term" list of the Cabinet Office's guidelines. CAVOC contains Agriculture Activity Ontology (AAO), Crop Vocabulary (CVO), and a machine-readable version of resources relevant to agriculture and food issued by Japanese government agencies.

(1) AAO

The main purposes of AAO are to describe data on activities in major crops such as rice, wheat, and tomato, and help share data among agricultural ICT systems. AAO defines the concepts of agricultural activities and the relationships between broader and narrower concepts, along with their properties and values, on the basis of Description Logics (Joo et al. 2016). For example, the concept of "Hand cutting" is described as "Agriculture_activity $\wedge$ $\forall$purpose.harvest $\wedge$ $\forall$act. reap $\wedge$ $\forall$equipment.sickle" (Fig. 1-2). Formalization by using Description Logics makes it possible to judge inconsistencies and subsumption among concepts, and thus enable logical inference. AAO contains 475 concepts of agricultural activities (2018.5.3, v. 2.01).

(2) CVO

In the food chain, data are key tools for demonstrating compliance with legal regulations and
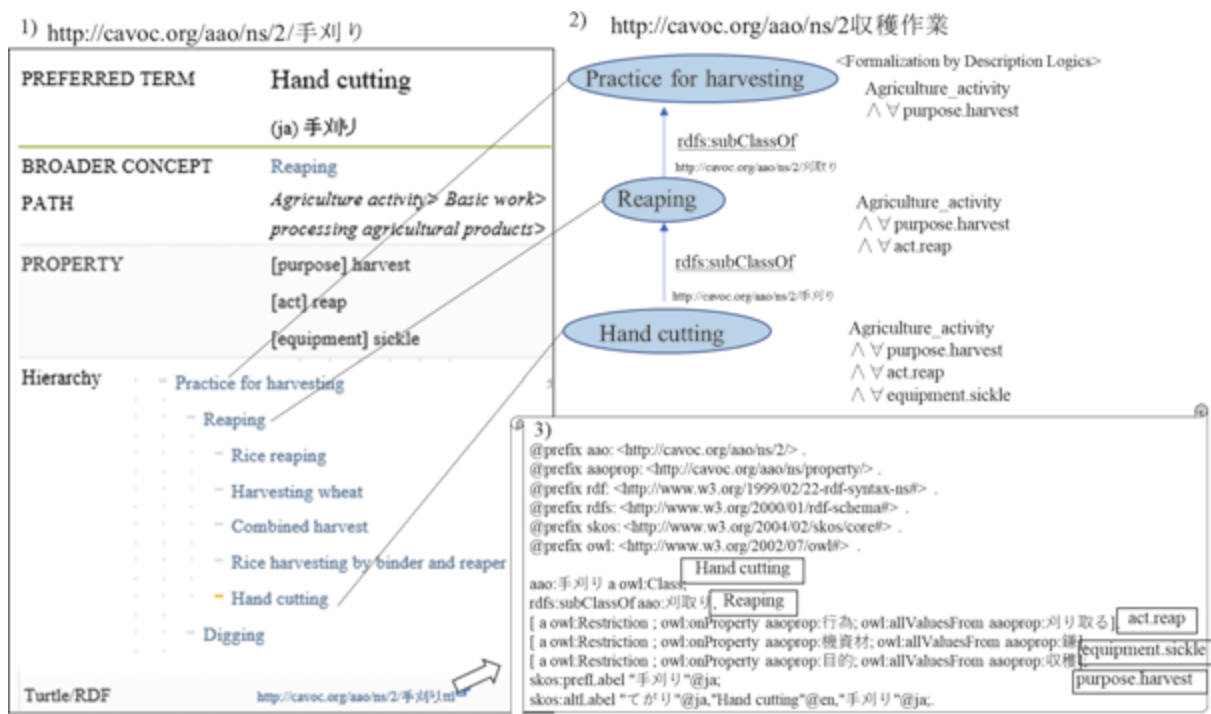

Fig. 1. Web Services based on Agriculture Activity Ontology
1) Example of uniform resource identifier (URI) for "Hand cutting" (http://cavoc.org/aao/ns/2/手刈り) in CAVOC
2) Example of a semantics graph in Agriculture Activity Ontology. Each concept is expressed by Description Logics. "Hand cutting" is described as:
"Agriculture_activity $\wedge$ $\forall$purpose.harvest $\wedge$ $\forall$act.reap $\wedge$ $\forall$equipment.sickle") and has a URI, as shown in 1).
3) The formula describing "Hand cutting" in Web Ontology Language (OWL)/Turtle format

thus raise societal expectations of food safety (CEMA 2017). The main purposes of building CVO were to share data between two sets of Japanese regulations on agricultural chemicals and help farmers comply with the regulations. One set of these regulations is the "Agricultural Chemical Use Reference" issued by the Ministry of Agriculture, Forestry and Fisheries, provided from the retrieval system at the Food and Agricultural Materials Inspection Center (http://www.acis.famic.go.jp/searchF/vtllm001.html). The other set is the "Agricultural Chemical Residue Reference" issued by the Ministry of Health, Labor and Welfare, provided from the retrieval system at The Japan Food Chemical Research Foundation (http://db.ffcr.or.jp/front/). Crop concepts in CVO are represented not only by organism names ("Cucumber") but also with added information, such as edible parts [Fig. 2-2, "Cucumber (leaves)," "Cucumber (flowers)," and "Cucumber (fruits)"] (Joo et al. 2018), because the acceptable limits of agricultural chemical use and residues differ among the different edible parts of a crop. Although the item names in the "Agricultural Chemical Use Reference" [Fig. 2-2, "Cucumber," "Cucumber (leaves)," "Cucumber (flowers)"] differ from those in the "Agricultural Chemical Residue Reference" [Fig. 2-2, "Cucumber,"

"Cucumber (leaves and flowers)"], data with item names that differ between both sets of regulations are integrated through CVO. Each crop concept in CVO contains the scientific name (Fig. 2, "Cucumis sativus"), the representative common name in both Japanese (Fig. 2, "キュウリ（果実)") and English (Fig. 2, "Cucumber (fruits)"), and synonyms and Japanese orthographic variations using three types of characters (hiragana, katakana, and kanji) and a mixture of these scripts (Fig. 2-2, "胡瓜（果実)"). CVO contains 1249 crop names (2018.5.2, v. 1.52).

The "Food Composition Tables" were published by the Ministry of Education, Culture, Sports, Science and Technology (https://fooddb.mext.go.jp/, accessed 30 July 2019). These tables provide basic data on food components and have been used widely to manage food services (e.g., school lunches, hospital food) and provide nutrition instructions (e.g., dietary restrictions, therapeutic diets), as well as advice regarding the daily home lives of those interested in nutrition and health (MEXT 2015). We have mapped crop concepts in CVO to food source organisms listed in the "Food Composition Tables" and to the standard terms included in the Cabinet Office's guidelines.
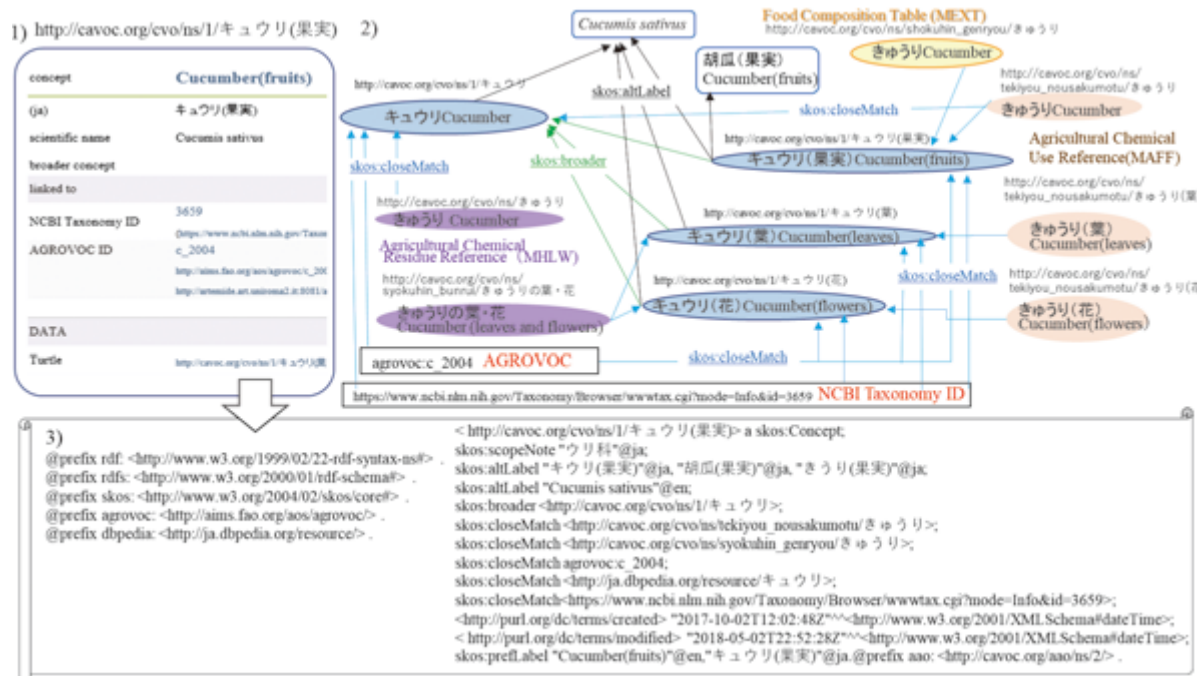

Fig. 2. Web Services based on Crop Vocabulary
    1) Example of uniform resource identifier (URI) for "Cucumber (fruits)" (http://cavoc.org/cvo/ns/1/キュウリ(果実))
    2) Example of a semantics graph in Crop Vocabulary (CVO). Each concept (circled) is linked to other concepts as per their relationships. ● CVO, ● Agricultural Chemical Residue Reference, ● Agricultural Chemical Use Reference, ● Food Composition Table, □ Scientific name, Japanese orthographic variations. Each concept has a URI.
    3) "Cucumber (fruits)" as described in SKOS format

2. Web service based on CAVOC

The CAVOC website (http://www.cavoc.org, accessed 30 July 2019) provides uniform resource identifiers by which the concepts in CAVOC are given (Figs. 1-1, 2-1). Its application programming interface allows external applications to look up representative common names in both AAO and CVO. The data in CAVOC can be downloaded in CSV, RDF/Turtle (AAO, Fig. 1-3), and SKOS/Turtle (CVO, Fig. 2-3) formats from the website. CAVOC is available as linked open data.

3. Applications

We have developed a tool for retrieving vegetable statistics, using CVO as a core vocabulary. Analysis of statistics on agricultural products throughout the food system is useful for helping new farmers choose the best crops based on crop productivity and profitability. CVO-mapped statistics can be easily accessed when the CVO crop name in the tool is selected. The example in Fig. 3 shows the statistical survey results of "Welsh onion" and vegetables representing narrower meanings, such as "Green welsh onion," "White welsh onion," and "Welsh onion harvested in summer" or "Welsh onion harvested in spring." The use of CAVOC for keyword expansion and indexing has also been proposed (Takezaki et al. 2017). Data annotated with different terms can be semantically integrated on the basis of CAVOC, thus enabling comparisons between data.

Other resources for data standards related to agriculture and food

Many vocabularies, ontologies, and taxonomies have been produced to represent and annotate agronomic data, but are scattered around the Web (Jonquet et al. 2018). To allow users to identify and select them for specific tasks, integrated resources and platforms have been developed within several international projects.

1. Resources for plant biology and crop breeding

Most efforts to secure interoperability using standards have been applied in the fields of plant biology and crop breeding. We introduce Crop Ontology and Planteome, as well as NCBI Taxonomy mapped to CVO.
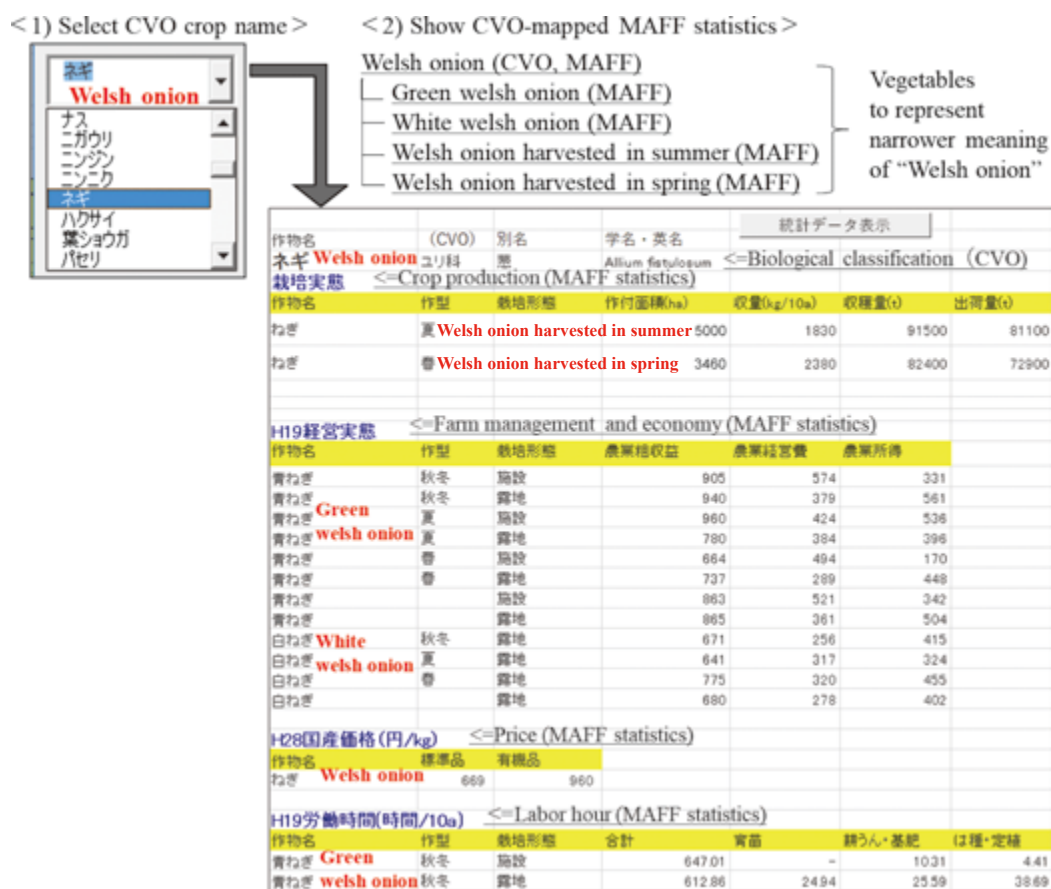

Fig. 3. MAFF statistical surveys of "Welsh onion" and vegetable terms representing its narrower meaning in our crop reference tool
CVO: Crop Vocabulary; MAFF: Ministry of Agriculture, Forestry and Fisheries.

(1) NCBI Taxonomy

The US National Center for Biotechnology Information (NCBI) Taxonomy database (http://www.ncbi.nlm.nih.gov/taxonomy, accessed 30 July 2019) is the standard nomenclature and classification repository for the International Nucleotide Sequence Database Collaboration, comprising the GenBank, EMBL, and DDBJ databases (Federhen 2012). The Taxonomy database is a central organizing hub for many of the resources at the NCBI, and provides a means of clustering elements within other domains of the NCBI website.

(2) Crop Ontology and Planteome

Crop Ontology provides several ontologies related to the anatomy, structure, and phenotype of crops; to trait measurement and methods; and to germplasm (https://www.cropontology.org/, accessed 30 July 2019). These ontologies are utilized in integrated breeding platforms and tools to annotate data on specific crops, thus making the annotated data discoverable, retrievable, and interoperable. Planteome (http://www.planteome.org/, accessed 30 July 2019) provides common ontologies for plants, including Plant Ontology and Plant Trait Ontology, Gene Ontology, Plane Experimental Conditions Ontology, Phenotypic Qualities Ontology, Chemical Entities of Biological Interest, Evidence and Conclusion Ontology, Planteome-NCBI Taxonomy, and crop-specific ontologies developed as part of the Crop Ontology Project (Cooper 2018). These ontologies map each other. Definitions can be highly specific by mapping crop-specific ontologies to common ontologies for plants. For example, "Sorghum stem length trait" in the Sorghum Ontology (one of the crop-specific ontologies) represents a concept that is more specific than the concept of "stem length" in the Plant Trait Ontology. Planteome also provides integrated data that are annotated with gene expression, traits, phenotypes, genomes, and germplasm across 95 plant taxa.

2. Broad scoped resources
(1) AGROVOC

AGROVOC is a large international vocabulary that covers domains such as food, nutrition, agriculture, fisheries, and forestry, and is published by the Food and Agriculture Organization of the United Nations (FAO 2019). It consists of 36,000+ concepts in up to 33 languages. AGROVOC is linked to 16 open datasets, including the US National Agricultural Library's Agricultural Thesaurus, DBpedia (https://wiki.dbpedia.org/, accessed 30 July 2019). Many datasets aligned to AGROVOC are available as linked open data. Consequently, a user can access these linked resources

as a single resource (Drury et al. 2019).

(2) AgroPortal and the VEST Registry

AgroPortal (http://agroportal.lirmm.fr/, accessed 30 July 2019) is a vocabulary and ontology repository that was developed to avoid the duplication of effort, and reuse previous community work (Jonquet et al. 2018). It provides a common platform to identify, host, and use 109 vocabularies and ontologies in the domains of agronomy, food, plant sciences, and biodiversity, with a focus on the use thereof for research data. The VEST Registry portal (http://aims.fao.org/vest-registry, accessed 30 July 2019) provides a list of 399 existing vocabularies (e.g., ontologies, thesauri, glossaries, schemas) with descriptions and a categorization of standards and links to the original websites, but not the content of the vocabularies or ontologies (Jonquet et al. 2018). It also provides a global map of existing vocabularies for the exchange of data in the domains of food, agriculture, and agricultural resources such as climate and the environment. The VEST Registry is linked with AgroPortal.

(3) Japanese resources

Not many semantics resources for use in data integration have been developed in the agricultural domain in Japan. An ontology management tool for the agricultural domain has been proposed for the purpose of registering, updating, and retrieving semantics data collected from existing resources (Nagai 2017), but is not currently open.

3. For data integration, the relationship between existing foreign standards and CAVOC

Resources for data integration in Japan must be enhanced by mapping to these foreign standards and extending them. To ensure international interoperability, both AAO and CVO link to AGROVOC, which is regarded as an international common vocabulary (Fig. 2-2, AGROVOC). Consequently, AAO and CVO may link to many datasets through AGROVOC. CVO also links to NCBI Taxonomy to ensure interoperability between data related to organisms and those related to crops as farm products (Fig. 2-2, NCBI Taxonomy ID). Crop concepts in CVO are represented not only by organism names ("Cucumber"), but also with the names of edible parts [Fig. 2-2, "Cucumber (leaves)," "Cucumber (flowers)," and "Cucumber (fruits)"]. Edible parts in CVO are described on the basis of plant morphology. It is possible that mapping of CVO to ontologies in Planteome, such as Plant Ontology and Plant Trait Ontology, which describe plant external structures, would help to integrate data related to plant traits with data on crops as farm products. Although an

analysis of both the VEST Registry and AgroPortal showed that nearly half of the resources in the agricultural domain are not available online, and that many are available only as PDF and fewer than half have application programming interfaces (APIs) (RDA Agrisemantics Working Group 2017), about one-third (98 resources) are machine-readable ontologies. Of these, 12 (including FoodOn) cover the category of food and human nutrition. FoodOn (http://foodon.org, accessed 30 July 2019) is a consortium-driven project intended to build a comprehensive and easily accessible global field-to-fork food ontology that accurately and consistently describes organisms as, for example, food sources, raw food products, or processed foods. The mapping of CVO to FoodOn would help to integrate data on crops as farm products and data on processed foods.

Information systems with semantic technologies

Agricultural semantics technologies have been used in knowledge-based systems for the integration of data from remote devices in decision-support systems, and in expert systems such as those for disease diagnosis (Drury et al. 2019). Ontology-based phenotyping information systems have been developed within some projects (Jayaraman et al. 2016, Neveu et al. 2019). In these systems, ontology and semantics graphs are used to unambiguously identify objects or traits in experiments and thus establish their relationships. Consequently, several kinds of experimental data are given in detail, along with specifically described information or metadata, and can be integrated and analyzed together. The construction of local ontologies by extending existing ontologies (Jayaraman et al. 2016) and mapping individual ontologies to existing standardized ontologies (Neveu et al. 2019) will ensure interoperability with external systems.

International collaborative efforts for agriculture standards

International collaborative efforts to create standards in agricultural semantics are underway, such as through mapping among semantic resources within international projects and by holding meetings among stakeholders, such as engineers and researchers. We need to join the international community in creating agricultural semantics standards by introducing the community to CAVOC, which is linked to AGROVOC and NCBI Taxonomy (Fig. 2-2). AGROVOC currently contains an inadequate number of concepts specific to Japanese agriculture, and more are needed. Concepts specific to Japanese agriculture must be proposed to the Food and Agriculture Organization (FAO) of the United Nations.

Future direction of our work, as determined by this review

To develop ontology-based information systems in Japan, various kinds of ontologies must be prepared to unambiguously define objects and their relationships. Not many semantics resources for standards have been developed in the domains of agriculture and food in Japan. Therefore, the extension of CAVOC coverage is important, but elaborate efforts are needed to build ontologies. We intend to join the international community on agricultural semantics standards so as to extend CAVOC coverage through the utilization of foreign ontologies.

Conclusion

Here, we have introduced the features of CAVOC and CAVOC-based services, and have reviewed other resources, information systems with semantic technologies, and international collaborative efforts to create standards in the domains of agriculture and food. We need to join the international community on agricultural semantics standards so as to extend CAVOC coverage and enhance the interoperability of Japanese data.

Acknowledgment

References

Cooper, L. et al. (2018) The Planteome database: an integrated resource for reference ontologies, plant genomics and phenomics. *Nucleic Acids Res*. D1168-D1180.

Drury B. et al. (2019) A survey of semantic web technology for agriculture. *Information Processing in Agriculture*, https://doi.org/10.1016/j.inpa.2019.02.001 (accessed 30 July 2019).

European Agricultural Machinery Association (CEMA) (2017) *Digital farming: what does it really mean?* https://www.cema-agri.org/images/publications/position-papers/CEMA_Digital_Farming_-_Agriculture_4.0_13_02_2017_0.pdf (accessed 30 July 2019).

Federhen, S. (2012) The NCBI Taxonomy database. *Nucleic Acids Res.* D136-D143.

Food and Agriculture Organization (FAO) of the United Nations (2019) *AGROVOC*. http://aims.fao.org/vest-registry/vocabularies/agrovoc (accessed 30 July 2019).

Jayaraman, P. P. et al. (2016) Internet of Things Platform for Smart Farming: Experiences and Lessons Learnt. *Sensors*, **16**, 1884.

Jonquet, C. et al. (2018) AgroPortal: A vocabulary and ontology repository for agronomy. *Computers and Electronics in Agriculture*, **144**, 126-143.

Joo, S. et al. (2016) Agriculture Activity Ontology: An Ontology for Core Vocabulary of Agriculture Activity. *In* Proceedings of the ISWC 2016 Posters & Demonstrations Track co-located with 15th International Semantic Web Conference, 1690.

Joo, S. et al. (2018) Building the Core Vocabulary of Crop Names to Integrate the Vocabularies by Multiple Government Agencies. *In* Semantic Technology: JIST 2018, Springer, Cham, 320-335.

Ministry of Education, Culture, Sports, Science and Technology (MEXT) (2015) *Standard Tables of Food Composition in Japan - 2015 - (7th Revised Edition)*. http://www.mext.go.jp/en/policy/science_technology/policy/title01/detail01/sdetail01/sdetail01/1385122.htm (accessed 30 July 2019).

Nagai, M. et al. (2017) Development of an Ontology Management Tool to Support Interoperability of Agricultural Information. *Agricultural Information Research*, **26**, 27-33.

Neveu P. et al. (2019) Dealing with multi-source and multi-scale information in plant phenomics: the ontology-driven Phenotyping Hybrid Information System. *New Physiologist*, **221**, 588-601.

Organization for Economic Co-operation and Development (OECD) (2019) Innovation, Agricultural Productivity and Sustainability in Japan. *OECD Food and Agricultural Reviews*, OECD Publishing, Paris, https://read.oecd-ilibrary.org/agriculture-and-food/innovation-agricultural-productivity-and-sustainability-in-japan_92b8dff7-en#page1 (accessed 30 July 2019).

RDA Agrisemantics Working Group (2017) *Landscaping the use of semantics to enhance the interoperability of agricultural data*. https://www.rd-alliance.org/system/files/documents/Deliverable1%20-%20Landscaping.pdf (accessed 30 July 2019).

Takezaki, A. et al. (2017) Effective use of datasets related to farm work based on Agriculture Activity Ontology. *Japan Journal of Farm Work Research*, **52**,187-189.

Wolfert, S. et al. (2014) A future internet collaboration platform for safe and healthy food from farm to fork. *In* 2014 Annual SRII Global Conference, 266-273.

Zeng, Marcia Lei (2019). Interoperability. *Knowledge Organization*, **46**, 122-146.