# Controlled vocabularies in metadata

**A case study on AGROVOC thesaurus use in FAO's large-scale metadata harvesting, integration and search system AGRIS**

Daniel Martini [1], Andrea Turbati[2] and Imma Subirats[3]

**Abstract:** Innovation in research and development relies on findability of information on prior work. For that purpose, FAO maintains AGRIS, an aggregator system making publication and dataset metadata distributed all across the globe searchable. Keyword metadata are a crucial factor for navigation and information retrieval in such systems. Thesauri like AGROVOC provide a means for standardized keyword assignment following best practices of subject indexing. A metadata analysis has been conducted with the objective of identifying common usage patterns and determining whether concept coverage in AGROVOC is adequate for annotation and indexing in AGRIS.

**Keywords:** metadata, text analysis, bibliometrics

## 1 Introduction and background

### 1.1 Metadata catalogs and the AGRIS system

Innovation in research and development and through education and capacity building relies to a significant extent on findability and accessibility of information on state-of-the-art and prior work like publications and datasets. For that purpose, computerized metadata registries and cataloguing systems have been put into place since the late 70s/early 80s that are meant to help users through dedicated search functionalities and interfaces to get at the information they are after. Such systems have been set up at organizations that might be called "knowledge organizations", for example, libraries, but also research institutions that produce or rely on larger corpora of publications and data sets in their day-to-day work. Having its roots in research information sharing, the rise of the World Wide Web almost naturally led to the networking of such catalog resources. Metadata standards like Dublin Core were developed [We95; DU20] accompanied by harvesting protocols like the

[1] Kuratorium für Technik und Bauwesen in der Landwirtschaft e.V., Arbeitsschwerpunkt Digitale Technologien, Bartningstraße 49, 64289 Darmstadt, d.martini@ktbl.de, https://orcid.org/0000-0002-6953-4524
[2] Food and Agriculture Organization of the United Nations, Viale delle Terme di Caracalla, 00153 Rome, Italy, andrea.turbati@fao.org
[3] Food and Agriculture Organization of the United Nations, Viale delle Terme di Caracalla, 00153 Rome, Italy, imma.subirats@fao.org

Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) [La02]. These developments enabled another layer to be put on top of catalogs and registries – aggregators that ingested metadata from several systems, providing unified search and access mechanisms across a distributed set of resources.

In the agricultural and food domain, one of the largest aggregators is the United Nations Food and Agriculture Organization's (FAO) AGRIS database [FAO24a]. It currently contains around 14 million metadata records on publications in 114 different languages. These are gathered from 1267 data providers from 161 countries all around the globe by either harvesting through standard protocols like OAI-PMH or leveraging other interfaces for metadata submission [FAO24b]. Metadata can be provided in different standardized formats. AGRIS recommends a minimum set of fields that should be present to achieve a certain quality level [FAO24c] and to that end also provides its own metadata application profile AGRIS AP [Sa05]. Among the data providers are, for example, agricultural and scientific libraries, but also research organizations, and altogether, they form the AGRIS network. A searchable registry provides more detailed information about participating providers [FAO24d].

Technically, metadata are stored in a MongoDB document store and are indexed for search using ElasticSearch. Apart from an interactive interface (accessible through [FAO24b]), data is provided as open data for download [FAO24e]. As a single download file would be prohibitively large, the open data set consists of an entry point XML file conforming to the W3C's Data Catalog (DCAT) vocabulary [Al24] and the Dublin Core Metadata Terms Specification [DU20] and that root entry file links to downloadable sub-catalogs for each provider. It is, however, important to be aware that the open data set is only a subset of AGRIS as some providers have not (yet) agreed to their data being shared under a license that is permissive enough.

### 1.2 Keywords, controlled vocabularies and AGROVOC

A typical key element of metadata are keywords. They play a crucial role, for example for grouping or classification of records, and are thus an enabler for implementing search functionalities like faceted search or recommender engines to suggest related material to users. Ideally, the keyword combination of a metadata record should provide a kind of "fingerprint" of that record, allowing quick identification of and limitation to the set of records of relevance to the user. Library science has a long-standing experience in systematic keyword assignment – also known as subject indexing. Controlled vocabularies and thesauri provide sets of terms that prevent metadata systems from having to deal with ambiguities of natural language and that can be organized, for example, in hierarchies to facilitate navigation and exploration. Standards have been developed that provide best practices and support and guide the implementation of cataloging as well as vocabulary maintenance systems [ISO11].

FAO's AGROVOC thesaurus is such a structured term set that was built for indexing purposes [FAO24f; Su22] covering all of FAO's topics of interest. In 1982, its first edition was published as a book, later on moving towards distribution over the internet using web technologies. By now, it is provided as a linked open data set based on semantic web standards like the Resource Description Framework (RDF) and the Simple Knowledge Organization System (SKOS) [MB09]. The core class of the SKOS specification is a concept that is defined as a "unit of thought" within the domain of discourse that the knowledge organization system should cover. Each such concept is globally uniquely identified using a (resolvable) Uniform Resource Identifier (URI) and can have preferred labels in different languages as well as alternative labels (synonyms). Language tags – as usual in RDF data sets – denote the respective label language. Concepts can be related to each other using relational predicates like 'broader' or 'narrower' but also semantically richer relations based on more elaborate ontologies. Figure 1 illustrates this model of the SKOS standard based on an example from AGROVOC, the concept with the URI http://aims.fao.org/aos/agrovoc/c_12151 that represents the 'lettuces'. The figure also highlights an accompaniment to AGROVOC: the AgrOntology that allows for modeling domain-specific relations within the thesaurus using a set of around 160 additional properties. In Figure 1, examples for these properties are agront:isProducedBy and agront:isUsedIn.

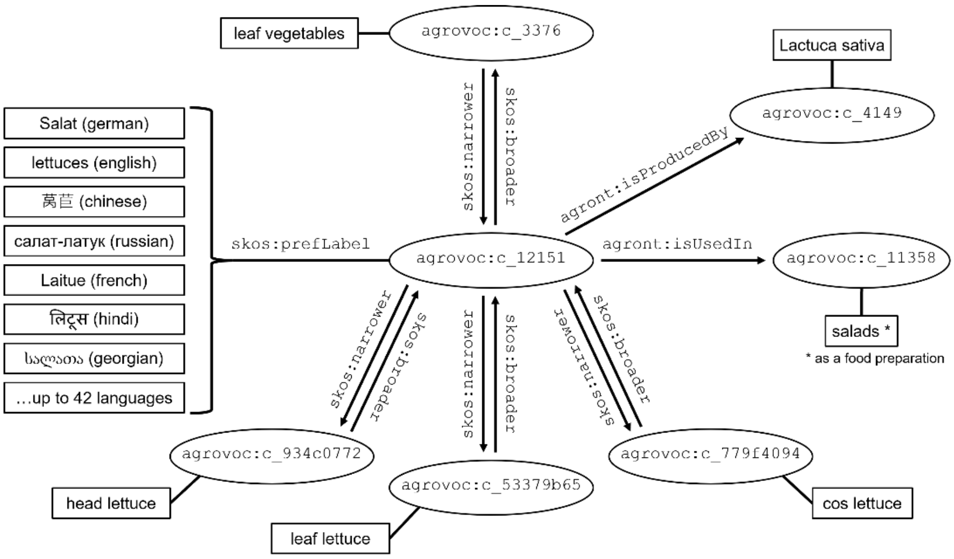


Fig. 1: SKOS model as used in AGROVOC, own illustration according to [MK22]. Text in ellipses represents concept URIs. Note that the shorter, prefixed form is used here. The string agrovoc: stands for http://aims.fao.org/aos/agrovoc/

At the top level of the hierarchy, generic, abstract concepts can be found like for example "activities", "objects", "organisms", "stages", "technology". Moving down the tree leads to more concrete concepts, for example "seed dressing" (in "activities") or "Bos taurus" (in "organisms"). AGROVOC currently follows a monthly release schedule, and its latest published version contains 41.163 concepts with over 1 million labels assigned to these concepts in up to 42 languages. It can be browsed interactively using a SKOSMOS interface [FAO24g; Su15] but aligned with best practices for linked data, it is also provided in a machine-actionable manner – concept URIs resolve to a ReSTful web service, a SPARQL interface [FAO24h; Ha13] is available for querying, and it can be downloaded in full in several standardized formats from FAO's data catalog [FAO24i].

### 1.3 Challenges and study purpose

One of the key principles of linked data and the semantic web is to "use URIs as names for things" [Be06]. The idea behind this is that it provides a globally unique identification for things and, if, additionally, HTTP URIs are used, these can be dereferentiated and agents can look up and use data that might be retrievable at these URIs. This mechanism corresponds roughly to symbolic names with global scope in the theoretical framework of formal and programming languages and it allows for lazy evaluation that is crucial for dealing with infinite data structures[4]. The benefit and need of globally unique and resolvable identifiers for data management purposes and good data stewardship has also been recognized within the framework of the FAIR principles for research data management [Wi16] – namely the F1 and A1 principles. The assumption that the benefit of using globally unique identification for interoperability and thereby weaving a world-spanning, machine-actionable knowledge graph would be obvious enough that metadata management systems would start using them directly in their datasets quickly and as broadly as possible was, however, probably overly optimistic. Although some larger organizations that curate metadata have moved towards linked data technologies – among them, for example, the German National Library (DNB) [DNB24] – the general uptake has been slow. If dealing with metadata that is gathered from distributed systems located in different countries you can therefore not rely on the presence of globally unique concept identifiers from controlled vocabularies for keywords. Instead, one has to be prepared to deal with keywords being provided as free text strings in a variety of different languages and with varying quality in terms of best subject indexing practices.

To gain a better understanding of keyword assignment practices and their development in different countries and institutions, a study on the AGRIS metadata set was conducted,

[4] Lazy evaluation commonly refers to a mechanism by which expressions or data are not immediately evaluated but rather machines decide on their own when results of evaluation are needed and only then calculations or retrieval are done – lazy evaluation is prominent in functional programming languages and enables dealing with virtually infinite data structures. The equivalence in linked data is deciding whether to follow a link and evaluate the data found there or not – within the virtually infinite World Wide Web.

specifically with the objective of assessing the usage of the AGROVOC thesaurus, to identify potential for improvement of its services and capacity building activities and to keep it up-to-date with the latest trends in scientific and technical literature related to food and agriculture, enhancing findability and interoperability. As a side result, the authors also expected insights into the temporal development of research topics in food and agriculture in different regions of the world. The following text presents the first part of the study that was dealing mainly with the former objective. The focus was on text-matching mechanisms in multilingual data sets, simple occurrence analysis and metadata quality challenges in metadata harvested from distributed sources. The outlook section provides an overview of work still in progress, including more elaborate bibliometric analysis, for example leveraging co-occurrence analysis and co-occurrence graphs for assessing subject relevance and its change over time.

## 2 Material and methods

### 2.1 Data sets used

From AGRIS, the full set of metadata records ingested into the system during the time span from the beginning of the year 2000 up to February 2024, including the ones not available in the open data set, was used for the analysis. For AGROVOC, the March 2024 released RDF data was used.

Keyword-label-matching data output was prepared as described in Section 2.2. Keywords and accompanying occurrence data were then loaded into a PostgreSQL database for subsequent querying and sorting.

### 2.2 Matching algorithm

The AGRIS dataset contains only free text string keywords in metadata. Although the authors know from personal communication that quite a number of data providers actually use AGROVOC for subject indexing purposes, indexing is done by textual (usually preferred) labels and the concept identifiers are either not captured in storage or they are not exposed to metadata harvesting or export interfaces. Therefore, AGROVOC had to be “retro-fitted” to the keywords present in AGRIS. For that purpose, an algorithm had to be developed to match character string keywords given in the respective metadata fields to concept labels in AGROVOC to determine whether a matching concept could be found. Because different conventions are followed by different data providers regarding handling of character modifiers like accents or usage of singular and plural, the authors distinguished between exact matches, normalized exact matches that match exactly if modifiers are removed/normalized and modified exact matches, with singular/plural conversion. An additional challenge was the multilinguality of both data sets. Language tags on keyword fields are only partially available in harvested metadata in AGRIS, and determining keyword language automatically reliably is, in many cases, practically infeasible. Some keywords in AGRIS match several different concepts in AGROVOC due to the same character string having different meanings in different languages. Additionally, it was determined whether the match was on a preferred label or on an alternative label, and a number of occurrences of keywords were counted. A Java program was put together for doing the matches that wrote as an output a set of CSV files, each containing the data from records for a single country for further processing.

### 2.3 Processing and statistics

As another processing step, metadata records were classified according to the percentage of matching keywords using the following classes, and the number of records within these classes was counted:

- no keywords at all present in the record
- no matching keywords present in the record (0% matches)
- records with > 0% and < 25% matches
- records with ≥ 25% and < 50% matches
- records with ≥ 50 % and < 75% matches
- records with ≥ 75% and < 100% matches
- records with fully matching keywords (100% matches)

This classification has been done using a Java program on a per country/per ingestion/publication year basis. For further analysis and visualization, the classification data have been loaded into a data frame in GNU R arranged along the factors of country of origin, publication year, and year the record was ingested in AGRIS.

On country and year level, relative proportions of these classes were calculated in relation to the total number of records – in other words, the percentage of records that exhibited 0 matches, the percentage of records that contained up to 25% of matches, the percentage of records that contained up to 50% matches and so on. To potentially discover change patterns, plots of these percentages over time have been generated for each country. The latter statistics were calculated using GNU R, and diagrams were generated using the ggplot2 package.

## 3 Results

There were around 4 million distinct keywords present in the whole AGRIS dataset, of which 136951 had at least one match to an AGROVOC concept. There were only 1212 keywords that matched 2 or more different AGROVOC concepts. Table 1 provides a list of the ten keywords that have the most occurrences in records.

| Keyword | number of occurrences | concept match | country usage |
|---|---|---|---|
| plant genetic resources for food and agriculture | 894587 | no | 1 |
| Models | 452193 | yes | 91 |
| Humans | 334675 | yes | 21 |
| Temperature | 333006 | yes | 96 |
| biology (general) | 322112 | no | 2 |
| Genes | 308285 | yes | 82 |
| Metabolism | 256662 | yes | 76 |
| Animals | 243028 | yes | 77 |
| Chemistry | 240410 | yes | 52 |
| Agriculture | 206415 | yes | 118 |

Tab. 1: Top 10 keywords by number of occurrence in records in AGRIS. Concept match denotes whether there was a match with an AGROVOC concept, and country usage denotes in how many countries that keyword was used

Subtracting the number of matching keywords from the total number of keywords found in the dataset gives us a long tail of non-matching keywords. Exploration revealed a number of reasons why no matches were revealed using the matching algorithm described in 2.2:

Two examples occur in Table 1: the most used non-matching keyword was ‘plant genetic resources for food and agriculture’, which in controlled vocabulary terminology is a so-called compound concept consisting of multiple words. Further investigation of the source data revealed that this keyword came from a single large data provider in Germany. It was probably assigned in an automated fashion to a large batch of metadata records. Another typical example is disambiguation by attaching qualifiers in brackets, as illustrated by the keyword ‘biology (general)’ in Table 1. Different wording can also be a reason for not finding a match: ‘united states’ was a keyword in 158918 records. The concept is actually represented in AGROVOC using its unambiguous, official name as provided by the United Nations’ names of countries list as its label: ‘United States of America’.

Apart from correctly spelled and encoded keywords, also erroneous or problematic entries with regard to interoperability were found. In the following examples, ‘’ are used only to clarify where strings found as keywords started and ended:

- Terms with quotes, dashes, slashes, hashtag, percent, spaces or other special symbols in the beginning/end or in odd places: ‘"broilers"’, ‘"curvina" larvae’, ‘" reactivity’, ‘--- biology’, ‘- cereals’, ‘#plante’, ‘#fridaysforfuture’,

‘%eichhornia’, ‘%lactobacillus’, ‘a griculture familiale’, ‘a grarprodukt’, ‘a. polygama’

- Plain numbers, either identifiers of unknown meaning or years
- Combinations of numbers with terms: ‘902.1 helmut josef braun’, ‘88spinacia oleracea’
- Terms with obvious character encoding errors: ‘ã¡cido ãºrico’, ‘â€œzona da mataâ€’, ‘ãâ‰-conotoxins’
- Cryptic codes: ‘a01n33’, ‘2nap’, ‘af026492’
- Terms with HTML-Tags or parts thereof in them: ‘<i>avibacterium paragallinarum<’, ‘<b>clinoptilolite’, ‘<em>blastomyces<’
- keywords that came from a single keywords field and that have obviously not been split/distributed correctly to several keyword fields in metadata: ‘discipline development|talent team|cooperation network|multiple linear regression’, ‘dieta||diet||dieta||régime alimentaire’

Challenges connected with processing such metadata entries are discussed in Section 4.

Figure 2 shows results of record classification and the respective percentages for a few selected examples that can serve to illustrate some typical patterns observed that will also be discussed in Section 4. The diagram contains an evaluation only for records that had a year of publication set in the metadata that was between 2000 and 2024. Similar diagrams that contain data for all records ingested by AGRIS during that time, for all countries and also by year have been generated during the course of the analysis. Presenting all material here would, however, go beyond the scope of this article.

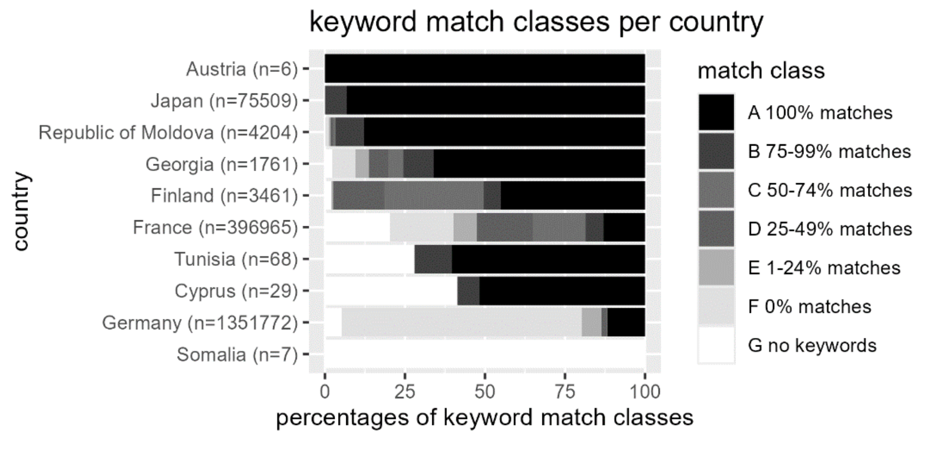


Fig. 2: Keyword match classes per country showcasing a few selected examples illustrating some characteristic observations made

## 4 Discussion

All in all, AGROVOC exhibited a good concept coverage of the most used keywords in the examined AGRIS dataset. Keywords with high occurrences that did not match could partly be ascribed a high relevance but also an inadequate wording choice as keyword terms. The term 'united states' that had a high occurrence in the dataset, for example, introduces ambiguity as it might also refer to the United States of Mexico, other formerly existing countries like the United States of Brazil, or other political entities like the United Arab States. Using complex compound concepts like 'plant genetic resources for food and agriculture' for keywords is a matter of debate. While they may increase specificity in information retrieval applications, they also make it significantly more difficult for users to find and combine the right terms to find what they are after. Therefore, it is often advisable to split such keywords into several terms. ISO25964-1 [ISO11] provides an in-depth discussion of the advantages and disadvantages of using complex compound concepts and factors to consider while deciding whether to introduce such concepts into a thesaurus and mentions, for example, frequency of use – given that today, large batches of records might get their keywords updated automatically, however, which can decrease usefulness of that criterion.

Also, the analysis shed a light on metadata quality. While issues like inadequate keyword splitting or remnants of HTML tags are easy to fix in an automated fashion, recognizing encoding errors or generalizing removal of unwanted characters can, however, be an immensely challenging task. Considering the case with numbers at the beginning of words, these can just as well be valid keywords, for example, in chemistry, like in 2,4-dinitrophenol, and at first glance, uninterpretable codes might refer to valid alphanumerical identification systems. Maintainers of harvesting and aggregating systems, therefore, have to carefully decide whether to touch the originally provided data at all – it is always a better option to enrich metadata instead of modifying it, but that comes at the expense of higher storage demands.

An aspect that Figure 2 highlights is that the number of records available for such an analysis is highly relevant. For Austria, all metadata records had a matching rate to AGROVOC of 100%. Austria, however, has only two data providers, and only two very small metadata ingestions took place during the time span considered in the data set – the total number of records (given in brackets after the country name in Figure 2) was therefore very low. At the other end of that scale is Somalia, which provided only 7 publication metadata records with a publication date between 2000 and 2024 – all of them without any keywords, demonstrating the devastating influence of political instability on knowledge-sharing activities. Japan, the Republic of Moldova, and Georgia provide examples of best practice controlled vocabulary use. Moldova and Georgia have an active AGROVOC user base with relevant institutional stakeholders engaged in capacity building and good metadata stewardship. The same holds true and is reflected in the data for Finland, although at a broader and more generic level, with the National Library of Finland actively developing tools like Skosmos [Su15] for controlled vocabulary provision and running initiatives like Finto [FT24] providing a number of thesauri and ontologies for subject indexing and information retrieval purposes. France represents an average, with a large number of records being provided that cover the whole range of matching classes. Tunisia and Cyprus might highlight a problem with large numbers of records with only single keywords being provided. In this case, the distribution of records across the matching classes is distorted towards the edges because either there is a match of that single keyword forcing the record to be classified as one with 100% matching keywords or none at all. Dependence of distribution on average numbers of keywords per record will, therefore, be a matter to be investigated in more detail in further analysis. Germany had a single metadata ingestion during the period considered that exhibited a high number of records that had keywords with no matches. This is mainly due to the mass assignment of ‘plant genetic resources for food and agriculture’ as a keyword; additionally metadata showed a high usage of cryptic codes as keywords. The issues highlight how trying to over-automate and over-systematize things can lead to the opposite of what was probably intended – mass assignments of the same keyword to a large number of records actually makes it more difficult to specifically find relevant records, and closed, undocumented and cryptic coding systems also add to the confusion rather than being helpful.

It also became clear that import patterns have to be taken into account in the analysis of the temporal development of concept usage – time series have not been shown here, but for some data providers, only single large imports were done, and it is impossible to determine when the metadata record was actually created – however, that would be the relevant information to analyze how AGROVOC content met demand at the time the subject indexing was done for the publication.

## 5 Outlook

Further work includes the application of statistical methods to work around the latter problem by analyzing larger sets of records and thereby determining the time a keyword was most probably most often assigned. Also, a closer examination of the occurrence numbers will probably deliver further insights into the relevance of AGROVOC concepts, concept groups, or relevance of subtrees of the concept hierarchy within the thesaurus. The application of further bibliometrical methods, especially co-occurrence analysis, should provide a map of the research landscape in agrifood by allowing for clusters of keywords/concepts to be determined [CLD21]. Currently, compilation and analysis of a more recent AGRIS data set has begun with improved timing information (record creation date, publication year etc.) with the additional objective of also developing tools for continuous analytics of incoming metadata with regard to potential AGROVOC extension.

## Bibliography


[Al24] Albertoni, R. et al. (eds.): Data Catalog Vocabulary (DCAT) – Version 3. World Wide Web Consortium, 2024, https://www.w3.org/TR/vocab-dcat-3/, Stand: 4.11.2024.

[CLD21] Chavalarias, D., Lobbe, Q. and Delanoë, A.: Draw me Science - multi-level and multi-scale reconstruction of knowledge dynamics with phylomemies. Scientometrics, 2021. https://nextcloud.iscpif.fr/index.php/s/XcLHCx9Zw89TeEP, Stand: 4.11.2024.

[Be06] Berners-Lee, T.: Design Issues Linked Data, 2006, https://www.w3.org/DesignIssues/LinkedData, Stand: 4.11.2024.

[DU20] DCMI Usage Board: DCMI Metadata Terms, 2020. https://www.dublincore.org/specifications/dublin-core/dcmi-terms/, Stand: 4.11.2024.

[DNB24] Deutsche Nationalbibliothek: Linked Data Service, https://www.dnb.de/lds, Stand: 4.11.2024.

[FAO24a] Food and Agriculture Organization of the United Nations: AGRIS. The International System for Agricultural Science and Technology – User guide. https://doi.org/10.4060/cc9993en, 2024.

[FAO24b] Food and Agriculture Organization of the United Nations: AGRIS – International System for Agricultural Science and Technology. https://agris.fao.org/en, Stand: 4.11.2024.

[FAO24c] Food and Agriculture Organization of the United Nations: AGRIS. The International System for Agricultural Science and Technology – Data provider guide, second edition. https://doi.org/10.4060/cc9512en, 2024.

[FAO24d] Food and Agriculture Organization of the United Nations: AGRIS Data Providers, https://www.fao.org/agris/data-providers, Stand: 4.11.2024.

[FAO24e] Food and Agriculture Organization of the United Nations: AGRIS Open Data Set (ODS), https://agris.fao.org/ods/AGRIS.ODS.xml, Stand: 4.11.2024.

[FAO24f] Food and Agriculture Organization of the United Nations: AGROVOC, https://www.fao.org/agrovoc/, Stand: 4.11.2024.

[FAO24g] Food and Agriculture Organization of the United Nations: AGROVOC multilingual Thesaurus – Skosmos Interface, https://agrovoc.fao.org/browse/agrovoc/en/, Stand: 4.11.2024.

[FAO24h] Food and Agriculture Organization of the United Nations: AGROVOC Thesaurus – SPARQL Service, https://agrovoc.fao.org/sparql, Stand: 4.11.2024.

[FAO24i] Food and Agriculture Organization of the United Nations: AGROVOC Thesaurus download in the FAO Data Repository, https://data.apps.fao.org/catalog/organization/agrovoc, Stand: 4.11.2024.

[FT24] Finish Thesaurus and Ontology Service, https://finto.fi/en/, Stand: 4.11.2024.

[Ha13] Harris, S. and Seaborne, A. (eds.): SPARQL 1.1 Query Language. World Wide Web Consortium, 2013, https://www.w3.org/TR/sparql-query/, Stand: 4.11.2024.

[ISO11] International Organization for Standardization: ISO25964-1: Information and Documentation – Thesauri and interoperability with other vocabularies – Part 1: Thesauri for information retrieval, 2011.

[La02] Lagoze, C. et al. (eds.): The Open Archives Initiative Protocol for Metadata Harvesting – Protocol Version 2.0, 2002. https://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm, Stand: 4.11.2024.

[MK22] Martini, D.; Kunisch, M.: Agricultural Data and Terminologies. In (Dörr, J. and Nachtmann, M.): Handbook Digital Farming – Digital Transformation for Sustainable Agriculture, Springer Verlag, Berlin, S. 153-162, 2022.

[MB09] Miles, A.; Bechhofer, S. (eds.): SKOS Simple Knowledge Organization System Reference. World Wide Web Consortium, 2009, https://www.w3.org/TR/skos-reference/, Stand: 4.11.2024.

[Sa05] Salokhe, G. et al.: The AGRIS Application Profile for the International Information System on Agricultural Sciences and Technology – Guidelines on Best Practices for Information Object Description, Food and Agriculture Organization of the United Nations, 2005.

[Su15] Suominen, O. et al.: Publishing SKOS vocabularies with Skosmos. Manuscript submitted for review, June 2015. https://skosmos.org/publishing-skos-vocabularies-with-skosmos.pdf, Stand: 4.11.2024.

[Su22] Subirats-Coll, I. et al.: AGROVOC: The linked data concept hub for food and agriculture. Computers and Electronics in Agriculture, vol. 196, 105965, Elsevier, https://doi.org/10.1016/j.compag.2020.105965, 2022.

[We95] Weibel, S.: Metadata: The Foundations of Resource Description, D-Lib Magazine, July 1995. http://www.dlib.org/dlib/July95/07weibel.html, Stand: 4.11.2024.

[Wi16] Wilkinson, M. et. al.: The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018. https://doi.org/10.1038/sdata.2016.18, 2016.