FSO: Food Safety Monitoring Ontology

Arif Yilmaz[1,*], Raghavendra Naidu[1] and Christopher Brewster[1,2]

[1]Institute of Data Science, Maastricht University, Paul-Henri Spaaklaan 1, 6229 GT Maastricht, The Netherlands
[2]Data Science Group, TNO, Soesterberg, The Netherlands

Abstract
Food safety is more than ever dependent on data collected about food samples by means of both laboratory analysis and, more recently, non-destructive sensors. The latter emerging technologies for food safety monitoring show great promise for ensuring the quality and suitability of food. Being non-destructive and quantitative as well as rapid and automated, they generate large amounts of data. However, because this is a relatively young field and given the large variety of measurement methods, types of sensors and devices available, so far there exist no data standards to enable data sharing and interoperability between data collections. Therefore, much of this data remains inaccessible in separate laboratories and research institutions limiting the community's ability to develop comprehensive data driven predictive analysis and monitoring models. In the context of the Horizon 2020 DiTECT project, we have developed semantically enabled software systems for standardized data management to enable data sharing and data analytics by stakeholders in the food safety sector. In this paper, we present DiTECT food safety ontology for covering food safety analysis technologies and methods for noninvasive and rapid assessment of safety and authenticity of food products. The ontology maximally reuses well-known concepts from widely used reference ontologies related to the domains of measurement, agriculture, food and biochemical analysis. Our expectation is that our ontology will become a key, publicly accessible resource for enabling the use of data science in food safety sector.

Keywords
Food safety, food safety monitoring, food analysis, ontology, CEUR-WS

1. Introduction

With the ever growing globalisation of the food system, food safety has become a significant challenge for modern societies [1]. There are a variety of hazards that can affect food integrity and thus impact human health, including adulteration, chemical contamination, and inappropriate storage and handling. The WHO estimates 600 million food-borne disease cases per year [2], and the EC's RASFF food alert platform reports an increasing number of food safety incidents in recent years [3]. Similarly for the US, a study by USDA ERC in 2021 reported the burden of foodborne ilnesses caused by pathogens on the US economy as $17.6 billion in year 2018 [4].

While there is a long history of careful laboratory based microbiological analysis of food samples, recent years have seen a move towards greater use of a variety of non-invasive/non-

SWAT4HCLS 2023: The 14th International Conference on Semantic Web Applications and Tools for Health Care and Life Sciences

*Corresponding author.
a.yilmaz@maastrichtuniversity.nl (A. Yilmaz); r.naidu@student.maastrichtuniversity.nl (R. Naidu); Christopher.Brewster@maastrichtuniversity.nl (C. Brewster)
0000-0001-6619-1469 (A. Yilmaz); 0000-0001-6631-1897 (R. Naidu); 0000-0001-6594-9178 (C. Brewster)

destructive analytical techniques for determining food authenticity, quality and safety. These methods generate data that must then be processed using statistical and other computational processes and thus have led to the adoption of data science methods and techniques to manage and analyse the data generated [5, 6]. However, as with most other scientific disciplines, a growing number of data sets naturally results in the growth of data silos and presents ongoing challenges to integrating data from multiple sources so as to develop systematic, high performance models.

The DiTECT project [1] funded by the EC and China aims to develop integrated framework for real-time detection, assessment, and mitigation of biological, chemical and environmental contaminants throughout the food supply chain. The project has two major axes. First, to collect a large set of food sample data, derived from a variety of sensor types, across four major food chains (maize/corn, beef, poultry, fish) and from multiple testing laboratories that are partners in the project so as to allow complex machine learning/deep learning models to be built predicting the safety of a given food sample. Second, to design a computational infrastructure that would (theoretically) operate across the supply chain, enabling samples to be taken, non-invasive, real-time tests to be made using the same variety of sensors and given the uploading of the sample data to receive a food quality assessment in response. It is in this context that the project is developing a data infrastructure that applies semantic technologies to ensure that data collected for both research and sampling purposes can be properly annotated and thus made reusable for future research and development purposes. In this paper, we present DiTECT Food Safety Ontology (FSO) which is designed to meet the data management needs of the application of data science to the food safety sector. The ontology is implemented in Web Ontology Language (OWL), available online in our repository[2].

This paper is structured as follows: This section motivates the need to adopt semantic technologies and thus to adopt appropriate tools such as food and food safety specific ontologies. The next section presents background on the challenges faced by the food safety sector and the growing importance of data science and AI techniques for this sector. Section 3 presents related work on ontologies for the agrifood sector and other areas that provide a foundation for the FSO. Section 4.1 presents our ontology building methodology, requirement analysis, competency questions, core concepts. In Section 5, we present and evaluate the ontology in detail, followed by a case study for use of FSO in processing a food sample and generating analysis results from a laboratory. This is followed by a brief discussion and conclusion describing planned future work.

2. Background

Monitoring and ensuring food safety and quality represents one of the biggest challenges for the food industry. Consumers expect high quality, safe and nutritious food from food retailers and food suppliers. Traditional methods of food safety inspection follow standardised regulations and are costly, time consuming, destructive of the food, and most importantly only retrospective. This consequently precludes real time intervention in the product's life cycle and

[1]https://ditect.eu/
[2]https://github.com/ditect-eu/ontology/

in the distribution network [6]. Food products are usually unsafe due to microbial pathogens or chemical contaminants and there has been a growth of food safety incidents in recent years due to variety of factors. These include changes in food processing methods, increases in globalisation and international trade, changing consumption patterns and possibly also increased consumer awareness. Governments around the world have established standardised regulatory inspections and sampling regimes which are largely based on ensuring processes followed are correct and the finished product is tested. As [6] points out, this approach is time consuming, costly, and destructive to the tested product. As a result, there is currently a move towards the use of non-invasive/non-destructive testing techniques using sensors based on a variety of disciplines including proteomics and other omics disciplines , gas chromatography, mass spectrometry and Fourier transform infrared spectroscopy. These analytical and high-throughput platforms generate massive amounts of data which need to be stored, appropriately annotated, and shared with other researchers to allow the creation of suitable signal processing and machine learning models to make practical use of the data in food testing contexts.

Given this background, the DiTECT project is building a prototype data management infrastructure. As is common in most areas of both theoretical and applied research, a variety of data silos naturally spring up, and in the case of food safety this means that each laboratory will use somewhat different parameters in setting up instruments, and once data is extracted from an instrument is most likely to store it in differently structured or differently interpreted formats. Even if Excel or CSV is used as a file format, the meaning of each column or row is highly context dependent. Consequently, a formal vocabulary or ontology is needed, as this provides convenient modelling structure for data annotation, data sharing and knowledge representation.

There are two aspects that a food safety ontology needs to address. One aspect concerns the food supply chain and the location along the supply chain that a food sample is taken, for example "on the farm", "at the abattoir", "in the grain mill" etc. This is not just a matter of location or supply chain step obviously but also the environmental context and any other parameters which might impact food safety. Steps in supply chains vary considerably between different types of food products (for example the presence or absence of a cold chain). The other aspect is the nature of the processing or instrument used to produce a test result through observable properties and scientific variables. At a specific step in a supply chain a large variety of instruments can be used and each produce different types of data. Furthermore, the data produced by an instrument is usually processed following one or more analytical methods (see [6] for a detailed catalogue). This paper and the ontology presented focus on the second aspect, i.e., the analysis and data associated of a specific food sample and these requirements are further described in Section 4.1 below.

3. Related Work

There exist a large number of ontologies related to, on the one hand agrifood, and on the other related relevant concepts in areas such as chemistry, scientific units and sensors. Here, we mention only the most important that have influenced our work or whose classes we directly inherited from. Our review of the literature has brought to light no dedicated ontology for food safety data and motivated us to develop and publish such a resource.

Agrifood related ontologies: There is a considerable body of work building ontologies for the food and agriculture domain which has gone hand in hand with the development of Linked Data (and "Linked Open Data") in the agri-food domain. The Agroportal lists over 140 ontologies [7]. The major effort here has been AGROVOC developed by the FAO and maintained by a network of institutes around the world. It is today the most comprehensive multilingual thesaurus and vocabulary for agriculture. Other recent work in this area has also focused on developing ontologies for sharing of research data including the Crop ontology initiative, the Agronomy Ontology (AgrO), and the Plant Trait Ontology (TO) supported by CGIAR. FOODON integrates a number of existing ontologies, but its focus seems to be again on research data, although its ambition is to provide a mechanism for data integration across the food system. Considerable efforts have been put into extending and integrating the FOODON ontology with various other ontologies extending its utility to areas such as nutrition and integrate it with the Foodex2 standard from EFSA. Most work on ontologies for the agrifood domain has up to now mostly been targeted towards the clear definition of domain concepts and terms in the form of a vocabulary for the annotation of research publications or research data sets.

Other relevant ontologies: There exist a number of other relevant ontologies which we have reused as much as possible, as discussed below. The most important is the well-known Semantic Sensor Network (SSN) ontology which describes sensors, observations, and related concepts and together with its successor the Sensor, Observation, Sample and Actuator ontology (SOSA) provides a reference ontology that defines classes such as sensors, devices, and samples.

4. Developing the Food Safety Monitoring Ontology

4.1. Methodology

In this section, we explain our ontology development approach based on standard methodologies used by the community [8]. **1. Determining scope, domain, and competency questions**. We relied on existing DiTECT project materials and the wider food safety literature, as well as interviews with DiTECT project partners to determine the ontology scope and to identify relevant competency questions. A subset of competency questions we identified is as follows (others are avilable in the DiTECT technical report):

1. *From what location was the sample taken?*
2. *Which sample types or features of interest are analysed for food safety?*
3. *What analysis results are provided after inspection of samples?*
4. *Which sensor types are utilised in analysis or monitoring of samples?*

2. Reusing relevant ontologies. In order to maximise interoperability, we reused as much as possible existing and publicly available ontologies. We queried the identified concepts on online search tools and ontology registries such as OLS [9], Ontobee [10], and Agroportal[7]. We filtered ontologies based on their relevance to the food safety domain, how widely used they are, coverage, semantic consistency, acceptance, and re-usability (particularly online availability in standard languages such as OWL). The ontologies selected are shown in Table 1 as well as the kind of classes reused.

Table 1
List of reused ontologies in food safety ontology.

Name	Domain	Purpose
SSN/SOSA	Sensor	sensor, device, sample etc.
CHMO	Chemical methods	analysis, process etc.
SAREF	Sensor	object properties
AGROVOC	Food	different agriculture domains
NCIT	Bio-Medical	latitude, longitude, location etc.
CHEBI	Chemistry	chemical compounds
Units of Measure (OM)	Mathematical Units	wavelength
SIO	Generic	mean and standard deviation

3. Checking the conceptual coherence of classes. Identified concepts are checked for completeness and consistency. For instance, in our scenarios, experiments are performed on sensors such as imaging sensors, and most devices contains one sensor, such as a multi-spectral imaging sensor. Some devices, such as e-nose may contain multiple sensors. Therefore, we defined devices and sensors as different classes. **4. Mapping and defining the class hierarchy**. Here, concepts and terms identified previously were mapped to specific ontology classes. We manually mapped each concept to classes in the selected ontologies to ensure the reusability and consistency of the FSO. The FSO was developed using the Protégé ontology editor version 5.5.0 [11]. We started with the most general classes and also working bottom-up from specific classes needed for certain analyses or instruments. **5. Verification of the ontology**. The implemented ontology was verified for logical inconsistencies in the classes and properties. Validation was undertaken to determine if the ontology could represent real world data. Further human validation was undertaken by organising six sessions with different researchers in the DiTECT project to validate the concepts, identify new concepts, enrich the ontology, and complete relations between classes.

4.2. FSO Description

Figure 1 shows part of the FSO related to classes for modelling food samples, produced data, results of the analyses on the samples. The whole ontology is available in our repository noted above.

Sample, SampleType, Dataset: The Sample class is derived from the SOSA ontology. It describes the entity on which the analysis is performed. The measurements performed on related samples are collected in a dataset. The samples are collected from a specific step in the supply. SampleType class holds the information on the type of sample collected. Location class keeps the information regarding the location from where the sample was collected such as a domain, or analysed such as a Laboratory. The Dataset class, derived from NCIT, consists of a set of measurements concerning a single food sample.

Feature of Interest, Property, Measurement, Unit of Measure: Figure 1 shows the structure of classes related to the measurement of a property. The FeatureOfInterest class is reused from SOSA which describes the entity for which a measurement is made. FeatureOfInterest is related to Property that is to be measured and Measurement that has information regarding

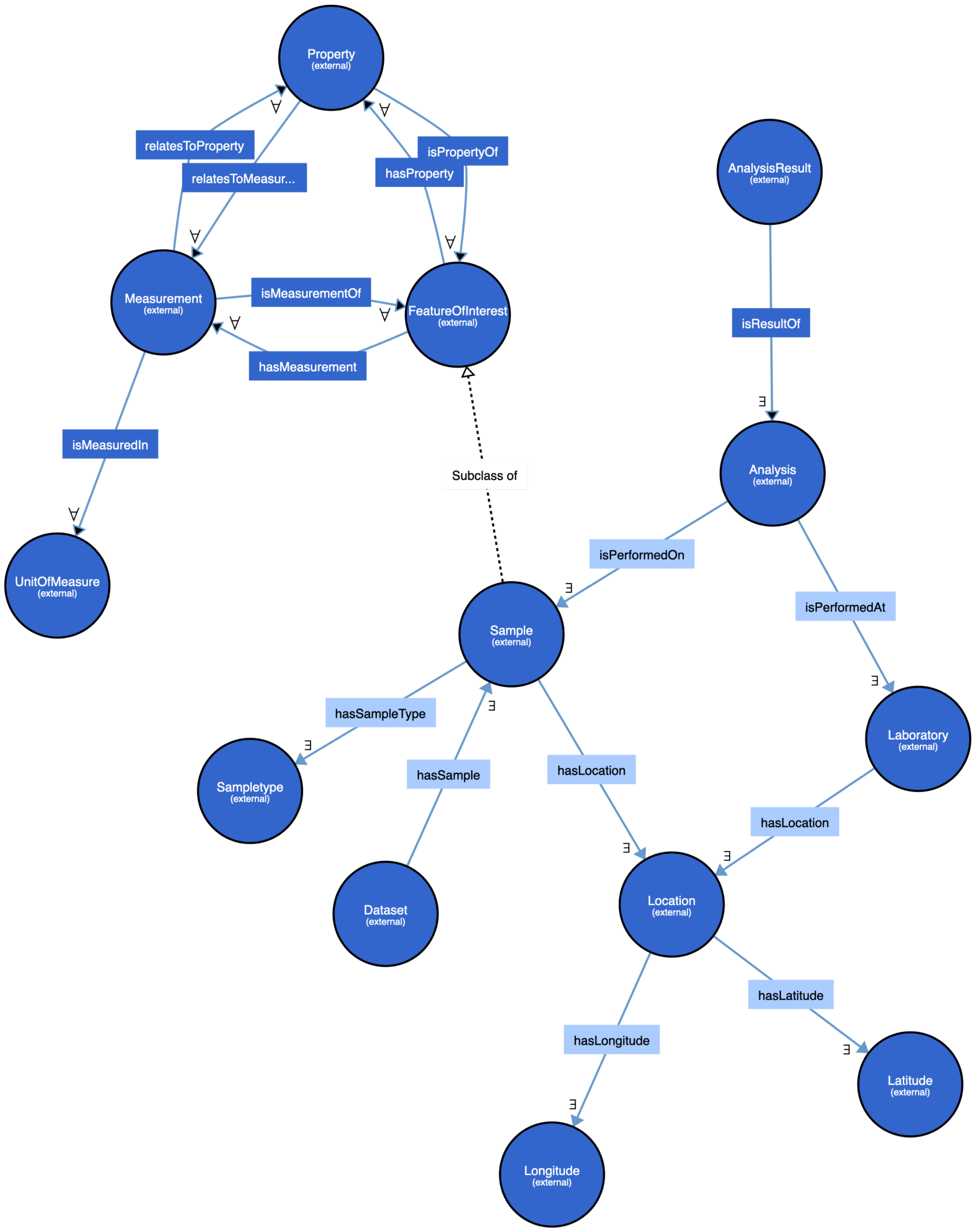

Figure 1: Food safety concepts and their relations. A sample is analyzed for a feature of interest. It is analyzed in a laboratory. Analsis results are produced. A dataset associated to a sample is generated as a file. Image generated with VOWL [12]

measurement details. The Property class is derived from the SSN ontology, and it describes the quality of a FeatureOfInterest e.g. carbon dixoide concentration. A property is measured by a Device. The UnitOfMeasure class is reused from the NCIT ontology, which basically defines the standard for measurement of a quantity. The Measurement class is derived from SOSA ontology. It defines the measured value over a property. The chemical compound classes are derived from the CHEBI ontology. Imaging sensors such as FTIR, NIR, UV make measurements in terms of Wavelength which is reused from the Units Of Measure Ontology. Classes for Mean and Standard Deviation measurements for the MSI sensor are reused from SIO ontology.

Laboratory, Analysis, Analysis Result: Figure 1 shows the relations of these classes. The Laboratory class is derived from the NCIT ontology. A Laboratory conducts an Analysis on a Sample which is collected from a step in the supply chain such as the SlaughterHouse. The class Analysis is derived from NCIT ontology and refers to the process of analysing the different samples obtained. It has two object properties, isPerformedAt relates to the class Laboratory and has information on the laboratory in which analysis was conducted, whereas isPerformedOn relates to the class Sample which has details on the sample on which analysis was performed. AnalysisResult class is derived from SOSA. It has one object property isResultOf which relates to class Analysis.

Location: The Location class is derived from NCIT ontology. It describes the geographical coordinates as Latitute and Longitude and address of an entity. Laboratory, A step in the supply chain, such as a SlaughterHouse also has Location properties.

5. Evaluation of FSO

The evaluation of an ontology is essential prior to deployment in semantically enabled software systems [13]. Our approach to evaluate the ontology was to validate it with real world data provided by project participants. The food safety sector does not yet have standardised forms for requesting a food sample analysis, so building on the expertise in the DiTECT project, we designed a semi-structured Analysis Request Form in Excel. This was subsequently filled in by a variety of project participants thereby providing a sample data set of instances. This enabled an evaluation of the ontology testing to what extent it was able to capture or correctly annotate the data (attributes/values) found in these filled-out forms. This section presents an evaluation of the FSO for two real world use cases concerning handling of a food sample from its collection to assessment of food safety analysis reports.

5.1. Case 1: Food Sample Analysis Request

The chicken fillet samples may be collected from various stages in the supply chain such as *Poultry, Slaughterhouse, Distributor and Processor*. Each stage has a unique location which is defined using *Latitute, Longitude* and *Address*. These features are necessary for further semantic reasoning and spatial and temporal analyses. Figure 2 shows a sample food safety laboratory analysis request form that is produced for the chicken fillets accepted by *Laboratory*. Initially, uniques identifiers are assigned to the analysis instance and the sample instance. Type of analysis such as food *Spoilage Analysis* or *Quality Assessment* is filled in the request form to

FOOD SAFETY LABORATORY ANALYSIS REQUEST FORM			
Date	22/8/2022		
Analysis	PoultryAnalysis-1	**Laboratory Name**	MicroLab-ML1
Sample Name	PA-1-ChickenFill etSample	**Address**	Paul-Henri Spaaklaan 1 (PHS1) 6229EN, Maastricht, The Netherlands
Sample Id	PA-1-00011	**Location**	*38.11257,* *23.27307*
Sample type	Chicken Breast Fillet	**e-mail**	*info*@microlab-ml1.nl
Analysis Types	Spoilage analysis	**Requested Measurements**	*Spectral image MSI, FTIR*

Figure 2: DiTECT laboratory measurement request form for spoilage analysis on chicken fillet food sample using MSI and FTIR methods.

specify the type of analysis. Subsequently, the required measurement types such as MSI and FTIR would be identified by experts.

5.2. Case 2: Food Sample Analysis Result

Figure 3 shows the generated form generated as a result of Multi-Spectral Imaging (MSI) measurements. The measurements are a representation of *Feature of Interest* of the analysis of the sample. Similarly, *Property* is the quality of the *Feature of Interest* that is being measured. In the analysis type of *DiTECT Sensor* and *Device* are set as Multi-spectral imaging (MSI) and Videometer respectively. *MeasurementType* is set as Reflectance. Reflectance *Measurement* readings are computed as mean and standard deviation for each wavelength and each item in the sample. Measurements are stored as separate csv files. The filename for the csv files are also registered in the form.

6. Conclusions

This paper has described the motivation for developing an ontology for the food safety sector and the principles we have followed. The ontology has been developed as part of the EC-funded DiTECT project in order to facilitate the sharing of food safety analysis data between food chain operators and food safety researchers. It will form a key technology in the DiTECT platform currently being built for food safety data management in both the EU and China.

DiTECT project utilises the ontology to create a common semantic model for food safety monitoring, analysis and knowledge discovery. The FSO encompasses all aspects of the DiTECT food safety monitoring platform, from data collection and performing analysis to reporting of

Laboratory Analysis Results

Food Sample MSI Measurement Analysis Result

Sample code number	PA-1-00011	**Measurement**	Reflectance
File Name	MSI-PA-1-00011.csv	**Format**	csv
Device	Videometer	**Sensor**	MSI

Id#	Mean 01	Mean 02	Mean 03	Mean 04	Mean 05	………	StdDev 01	StdDev 02	StdDev 03	StdDev 04	StdDev 05	……..
1	13.353	18.442	23.224	29.869	33.547	……	2.202	3.078	3.514	3.962	4.204	……

Figure 3: MSI measurement for chicken fillet spoilage analysis.In the analysis type of *DiTECT Sensor* and *Device* are shown as multi-spectral imaging (MSI) and Videometer and *MeasurementType* as Reflectance is set for each wavelength.

results. Reuse of existing ontologies enabled both faster implementation of the FSO and reduced the effort to discover and realise relevant concepts. We evaluated the DiTECT food safety monitoring ontology (FSO) through laboratory analysis use cases. The main focus of future work will be to extend the ontology to properly capture the steps along the food supply chain including different types of food producers and food processors. Our next future perspective is to publish FSO ontology and improve its usability by addressing principles from guidelines such as MIRO[14] and I-ADOPT [15].

6.0.1. Acknowledgements

This study is funded from the European Union's Horizon 2020 research and innovation program with the acronym "DiTECT" under grant agreement No 861915.

References

[1] D. R. Lineback, A. Pirlet, J.-W. Van Der Kamp, R. Wood, Globalization, food safety issues & role of international standards, Quality Assurance and Safety of Crops & Foods 1 (2009) 23–27. doi:10.1111/j.1757-837X.2009.00005.x.

[2] WHO, Food Safety - Fact Sheet, 2022. URL: https://www.who.int/news-room/fact-sheets/detail/food-safety.

[3] A. Kowalska, L. Manning, Using the rapid alert system for food and feed: Potential benefits and problems on data interpretation, Critical Reviews in Food Science and Nutrition 61 (2021) 906–919. doi:10.1080/10408398.2020.1747978.

[4] S. Hoffmann, Ahn, Jae-Wan, Updating Economic Burden of Foodborne Diseases Estimates for Inflation and Income Growth (2021) 33.

[5] A. I. Ropodi, E. Z. Panagou, G. J. E. Nychas, Data mining derived from food analyses using non-invasive/non-destructive analytical techniques; determination of food authenticity, quality & safety in tandem with computer science disciplines, Trends in Food Science

& Technology 50 (2016) 11–25. URL: http://www.sciencedirect.com/science/article/pii/S0924224415301424. doi:10.1016/j.tifs.2016.01.011.

[6] G.-J. Nychas, E. Sims, P. Tsakanikas, F. Mohareb, Data Science in the Food Industry, Annual Review of Biomedical Data Science 4 (2021) 341–367. doi:10.1146/annurev-biodatasci-020221-123602.

[7] C. Jonquet, A. Toulet, E. Arnaud, S. Aubin, E. Dzalé Yeumo, V. Emonet, J. Graybeal, M.-A. Laporte, M. A. Musen, V. Pesce, P. Larmande, AgroPortal: A vocabulary and ontology repository for agronomy, Computers and Electronics in Agriculture 144 (2018) 126–143. doi:10.1016/j.compag.2017.10.012.

[8] H. Panoutsopoulos, C. Brewster, S. Fountas, A semantic data model for a FAIR digital repository of heterogeneous agricultural digital objects, in: Integrated Food Ontology Workshop, Bolzano, 2021.

[9] S. Jupp, T. Burdett, C. Leroy, H. E. Parkinson, A new ontology lookup service at EMBL-EBI, in: J. Malone, R. Stevens, K. Forsberg, A. Splendiani (Eds.), Proceedings of the 8th Semantic Web Applications and Tools for Life Sciences International Conference, Cambridge UK, December 7-10, 2015, volume 1546 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2015, pp. 118–119. URL: http://ceur-ws.org/Vol-1546/paper_29.pdf.

[10] E. Ong, Z. Xiang, B. Zhao, Y. Liu, Y. Lin, J. Zheng, C. Mungall, M. Courtot, A. Ruttenberg, Y. He, Ontobee: A linked ontology data server to support ontology term dereferencing, linkage, query and integration, Nucleic Acids Research 45 (2017) D347–D352. doi:10.1093/nar/gkw918.

[11] M. A. Musen, The protégé project: a look back and a look forward, AI matters 1 (2015) 4–12.

[12] S. Lohmann, S. Negru, F. Haag, T. Ertl, Visualizing ontologies with VOWL, Semantic Web 7 (2016) 399–419. URL: http://dx.doi.org/10.3233/SW-150200. doi:10.3233/SW-150200.

[13] C. Brewster, H. Alani, S. Dasmahapatra, Y. Wilks, Data Driven Ontology Evaluation, in: Proceedings of the International Conference on Language Resources and Evaluation (LREC-04), Lisbon,Portugal, 2004.

[14] N. Matentzoglu, J. Malone, C. Mungall, R. Stevens, Miro: guidelines for minimum information for the reporting of an ontology, Journal of biomedical semantics 9 (2018) 1–13.

[15] B. Magagna, G. Moncoiffe, M. Stoica, A. Devaraju, A. Pamment, S. Schindler, R. Huber, The i-adopt interoperability framework: a proposal for fairer observable property descriptions, in: EGU General Assembly Conference Abstracts, 2021, pp. EGU21–13155.