Ontology-driven Data Management Design in Healthcare Domain: The ADCATER Experience

Leonardo Cocco[1], Paolo Fantozzi[2], Domenico Lembo[1,*], Umberto Nanni[1] and Federico Maria Scafoglieri[1,*]

[1]*Department of Computer, Control and Management Engineering, University of Rome la Sapienza, Rome 00185, Italy*
[2]*Department of Law, Economics, Politics, and Modern Languages, LUMSA University, Rome 00192, Italy*

Abstract

In this paper, we outline our experience in implementing the data management component of a data-intensive healthcare application within the `ADCATER` project (Advanced Digital Solutions for Professional Food and Nutrition Catering Service), where an ad hoc ontology, tailored to the project's domain, plays a crucial role in driving the system design. We will discuss the creation process of this ontology, its underlying building principles, and how it aids the development of a reconciled database useful to integrate and consolidate heterogeneous sources of information, vital for the proper running of solution at hand. Here, the ontology is essential for harmonizing vocabularies and ensuring the establishment of a schema devoid of inconsistencies. Finally, we will explore how Business Intelligence services, operating on the foundation of the Data Warehouse built upon the reconciled database, are seamlessly aligned with the crafted ontology.

Keywords

Ontology, Data Management Design, Healthcare, Business Intelligence

1. Introduction

The Advanced Digital Solutions for Professional Food and Nutrition Catering Service (`ADCATER`)[1] is an international consortium focused on assisting healthcare professionals, especially nutritionists, in measuring and managing hospitalized patients' dietary intake via a digital meal tracking solution. This healthcare application comprises two main components: a computer vision-based system responsible for of identifying and categorizing patients' food consumption from images taken before and after meals, and a data management component that integrates this information with data stored in various data sources. This new integrated data-layer, suitably enhanced with Business Intelligence (BI) techniques, enables automatic reporting for proactive patient care initiatives, forecasting and preventing possible outbreaks.

In this paper, we will focus on the data management component, which is designed based on an ontology specifically created for this purpose, called in the following `ADCATER` ontology.

Ontologies formally conceptualize domains of interest, providing a common vocabulary of classes, relationship between them and properties, emphasizing the sharing of knowledge and

[1]https://ictagrifood.eu/node/44630

the consensus about its representation [1]. Formalized through an interconnected semantic network of information units, and sometimes called also Knowledge Graphs [2], they are the backbone of Semantic Web [3], where the formalism for their specification has been concretely standardized in the Ontology Web Language (OWL), finding a spread of applications in different contexts. When ontologies are coupled with data [4, 5], they prove to be valuable allies for data management [6]. They allow for semantic enrichment of the data to which they are connected, enable forms of reasoning to enhance the information services and the quality of the data itself [7]. This setting is usually known in literature as Ontology-based Data Management (OBDM) [4, 8].

While OBDM is preferable for expressing the potential of the ontologies in data-intensive applications, especially those that also require data integration [9], in some contexts these approaches and their implementations are not mature enough to be used or there are some restrictions imposed by the working environment. This is precisely the case for the project discussed in this paper. The project requirements within ADCATER impose specific constraints. For example, the cloud technology stack defined for the solution in production must interface with the existing applications of the industrial partner, FoodFix[2], which are unable to accommodate OBDM implementations. Additionally, the data architecture comprises BI technologies, which poorly match OBDM. Furthermore, there is an inability to preserve the privacy of the stored users' data in the current OBDM systems, which is essential in healthcare settings.

While having these project restraints, the role of the ADCATER ontology lies in driving the design of the data management component, standardizing the structure and vocabulary of the built solution. This feature is crucial in large projects and also allows the implementation to be open to interfaces with other external systems in a more standardized and uniform way. Moreover as mentioned before, the ADCATER ontology through a semiautomatic approximation and translation is useful to generate a inconsistency-free reconciled schema integrating and semantically harmonizing different data sources. Here, the reconciled database essentially bridges the gap between OBDM and BI technologies, aligning them with ontology terminology and basing their multidimensional objects (a.k.a. Cubes) on it. In order to openly share the material, the ADCATER ontology can be downloaded from the following link https://tinyurl.com/4h95yvav.

The paper is structured as follows. Section 1 is this introduction. In Section 2 we talk about the source of knowledge to build the ADCATER ontology, the methodology and the tools used. In Section 3 we introduce the ADCATER ontology and its modules. In Section 4 we discuss the reconciled database and the BI services built starting from the ADCATER ontology. The conclusion and the future works are delineated in Section 5.

2. Knowledge sources, methodology and tools

The ADCATER ontology, developed within a multidisciplinary project, formalizes and integrates several domains, including biomedical/healthcare, with focus on patients' status from a feeding perspective, as well as nutrition and food related entities and processes.

For the formalization of the ontology discussed in this paper, our primary sources of knowledge to be conceptualized are derived from *(i)* medical questionnaires administered to patients,

---

[2]https://foodfixit.com/en/foodfix-the-right-solution-for-smart-food-management/

(ii) existing applications that handle medical data and *(iii)* insights gathered through interviews with domain experts.

  (i) **Medical Questionnaires**. These questionnaires, mainly paper-based, serve two primary purposes: those conducted upon admission provide valuable insights into the patient's medical history, while those administered during the hospitalization stay help in monitoring the progression of the illness. These contain essential details for conceptual design purposes, such as the information needed to calculate indicators to verify the patient's nutritional health. For example, among these, there are the patient's weight, her/his height and age, which translated into the Body Mass Index (BMI) parameter are useful to compute a pleotra of standard scores for the assessment of malnutrition (GLIM [10], SNAQ [11] etc.).

  (ii) **Existing Healthcare Applications**. FoodFix is the industrial partner of the `ADCATER` project whose main business is the catering management of hospitals. FoodFix's systems, and in particular their databases, although not in a terminologically standardized form and not arranged for data-analysis purposes but mainly for operational ones, contain useful information regarding food. Properly, they store fine-grain details concerning patients' meals, how they are made and the macro-nutrients composing them.

  (iii) **Interviews with Domain Experts**. Undoubtedly, the primary source of knowledge and guidance in refining an ontology lies with domain experts. This also holds for the `ADCATER` ontology, which draws upon insights provided by nutritionists, physicians, and experts in hospital settings. Each of them contributed to a specific part (module) of the `ADCATER` ontology, either through interviews—question-and-answer sessions focusing on different aspects of the ontology, or by expressing their requirements. Their contributions were then traced back to verify the proper conceptualization.

The creation of the ontology from a methodological point of view was accomplished through an iterative refinement approach, which occurred through specialization of concepts and relationships and by modularizing its parts. The vocabulary of the `ADCATER` ontology also follows a logic of sharing, and thus all names used for artifacts were terminologically accepted by all parties involved. To support all the creation steps and to facilitate the communication with non data modeling experts, we were supported by a formal visual language for defining ontologies called Graphol [12]. This language is the basis of a graphical tool called Eddy[3]. Such a tool makes it possible to create ontologies using all the expressiveness of OWL through a visual graph-like representation similar to UML and Entity-relationships diagrams. It allows to export the ontology in text format following Semantic Web standards. Eddy is also useful for checking possible inconsistencies at ontology building time. That is, through automatic reasoning services related to the ontologies formal language, it is possible to check immediately whether intensional inconsistency or cases of non-instantiable concepts or relations arise. To facilitate the ontology definition process, we utilized collaborative text documents shared among all project partners. These documents provide detailed descriptions of the technical choices made for the formalization of each concept, relationship, and attribute within the ontology. This approach enables us to track offline changes and identify any potential discrepancies

---

[3]https://github.com/obdasystems/eddy

3. `ADCATER` Ontology

The `ADCATER` ontology is formalized using several ontology design patterns [13]. We exploit the logical patterns *N-Ary Relation* and *Tree* appropriately suitably translated into the Graphol language. Note that, the main purpose of the ontology is to guide the design of the data-related components of the ADCTARE solution, and thus it is not intended to cover all domain aspects. Metrics detailing the size of the `ADCATER` Ontology, including the number of classes, relationships and attributes, are given in Table 3.

	Number of
Concepts	43
Relationships	28
Attributes	65
Axioms	461

3.1. Modules

The `ADCATER` ontology is split into four modules, each of which is connected to at least one of the others and named according to the aspects it covers. This division into modules facilitates the creation and refinement of the ontology, Moreover, it has been of great help in the interviews with the experts and to verify where the conceptualized knowledge comes from.

Figure 1 shows a general overview of the ontology highlighting the modules and the main concepts belonging to them.
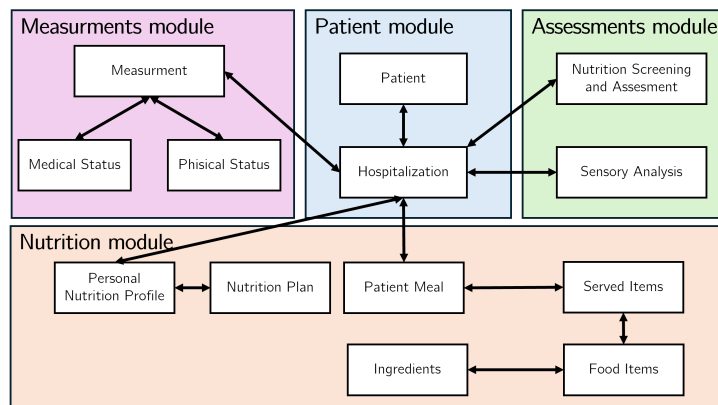
Figure 1: General overview of the ontology highlighting the modules and the main concepts belonging to them. Here, the rectangles representing the ontology concepts, and the arrows are ontology relationships that exist directly or indirectly (passing through other concepts or relationships) between them.

The four modules that define the `ADCATER` ontology are: *Patient*, *Assessments*, *Measurements*, and *Nutrition*. The pivotal concept that unites the modules is that of `Hospitalization`.

Figure 2 illustrates an excerpt of `ADCATER` ontology in Graphol, belonging to the Patient module, focusing on the `Hospitalization` concept. Here rectangles define concepts, diamonds denote relationships between them, and circles represent concept attributes. In the figure, the
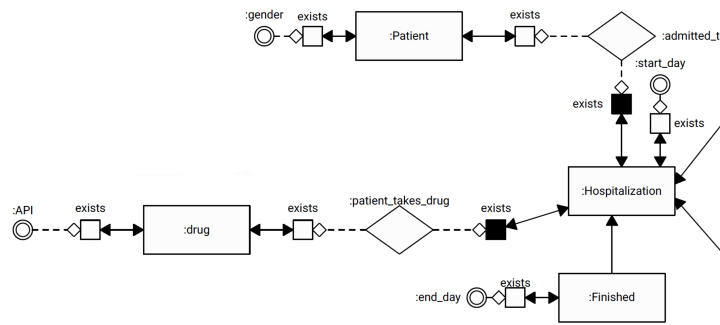
Figure 2: Excerpt in Graphol of the ADCATER Ontology, focusing on the Hospitalization concept.

concept `Hospitalization`, designed to represent the periods in which patient are admitted to the hospital, has the attribute `start_day` tracking the start date of them. Patients, represented by the class of the same name, have the attribute `gender`, identifying the patient's gender and is connected to `Hospitalization` with the relationship `admitted_to`. The ontology excerpt also models `finished` hospitalizations, and the drugs (identified by API, i.e., the Active Pharmaceutical Ingredient) administered to patients during their hospitalizations.

In the following we briefly discuss every modules.

Patient: This module deals with formalizing knowledge related to patients and the hospitals in which they are hospitalized. The main concept on which the module revolves is that of the patient (`Patient`). As discussed above, through the hospitalization concept, it is possible to track information about the period in which patients have been in an hospital, the hospital where they have been admitted, as well as their medical prescriptions.

Assessments: Patient assessments corresponds to clinical analyses, whose detail and outcomes are derived from medical questionnaires, and are used to compute scores for malnutrition indicators. Each type of assessment is represented by a concept, whose attributes are useful to calculate the indicators. The ontology also keeps track of the date of the calculation. This thus makes it possible to historicize information, which is crucial for temporal analysis. In this module, we also monitor the patient's appetite, along with tracking information about their sensory abilities, such as how they perceive taste and smells.

Measurements: This module is designed to represent the physical and psychological status of the patient. It contains key and historicized information on numerical indicators such as BMI, height etc., as well as mobility and muscle mass. This module also takes care of keeping track of the outcomes of medical teams, i.e. their diagnoses. This conceptualization, appropriately incorporated with the other modules, allows for a more complete view of the inpatient stay.

Nutrition: The nutrition module is concerned with food related information, for patient food intake modeling and monitoring. Here, the patient's nutritional profile is taken into account and the nutritional plan assigned to the patient is mapped out. Data intended to instantiate this module come from the computer vision component of the ADCATER solution, thus allowing for the actual food consumption to be tracked. The amount of food consumed by a patient is crucial information that the ontology models for comparison with prescribed food. The intake

of macronutrients can be easily obtained from the ingredients of the food items represented in the ontology and the quantity of food consumed.

3.2. Related Resources

We conclude this section discussing related resources (ontologies, standard vocabularies etc.) that have points in common with the `ADCATER` ontology.

We will focus on resources pertaining to the areas of biomedical/healthcare field (which we refer to simply as healthcare) and the food field as the `ADCATER` ontology does.

Healthcare ontologies. Ontologies, in the field of healthcare, describe the concepts of medical terminologies and the relation between them, thus, enabling the sharing of medical knowledge. Many medical ontologies are simply hierarchical vocabularies, in which the most general terms appear at the hierarchy top-levels and the terms become more specific down the hierarchy. The ontologies are typically quite big. The level of detail they reach is very high, and typically much higher than the level needed in `ADCATER`. We below mention some of them:

- SNOMED [14]: it is a family of terminological systems, which is around by more than 40 years. In particular, SNOMED CT (clinical terms) [15], which merges together the previous SNOMED RT [14] and Clinical Terms Version 3, is considered one of the most comprehensive, multilingual clinical healthcare terminology in the world. SNOMED CT provides several hierarchies of terms and includes Description Logics axioms. SNOMED CT establishes a vocabulary for electronic health records, including symptoms, diagnoses, medicines, etc.
- International Classification of Diseases (ICD) [16]: it is a nomenclature to classify diseases, injuries, and causes of death. It is maintained by the World Health Organization (WHO) and revised periodically. The current version is ICD-11 [17]. Some efforts have been made in the literature to identify aspects that ICD-11 has in common with SNOMED. Nonetheless, they remain to date different vocabularies.
- The National Cancer Institute (NCI) Thesaurus [18]: it is a description logic-based terminology, which is a part of the US National Cancer Institute Bioinformatics. It has been created to be used by NCI's researchers and the whole cancer community. It is designed to serve several purposes such as annotation, search, and retrieval of data, automated indexing, retrieving bibliography information, and linkage to heterogeneous resources.
- Medical Subject Headings (MeSH) [19]: it is a vocabulary specifically created for indexing journal articles and books in the life sciences. It is managed by the United States National Library of Medicine (NLM). It is also used to classify diseases studied by trials included in the ClinicalTrials.org.

Food ontologies. The definition of food thesauri, vocabularies and ontologies is more recent with respect to the analogous effort done in the medical domain. Nonetheless, several resources do exist that aim at promoting the standardization of terminology to be used to describe various aspects of the Food domain, from the names of animals, plants, and fungi that can be used as food for humans or domesticated animals, to prepared food and related products and processes. Below we mention some initiatives and available controlled vocabularies:

- FoodWiki [20]: it is a Mobile Safe Food Consumption System based on the Food Ontology Knowledge Base (FOKB), an OWL ontology that describes various kinds of food, accompanied with nutritional values, and recommendation about daily assumption. FOKB is structured in four subsections: person, disease, product, and food ingredients/compounds.
- AGROVOC [21]: it is a thesaurus providing terms in various languages to describe data in the agriculture, fishing, forestry, and food domains.
- FoodOn [22]: it is a comprehensive food ontology belonging to the open source OBO Foundry registry of ontologies for interoperable life science. It was originally based on LanguaL, a food indexing system for the description of food source plant and animal organisms, food preservation, cooking, packaging, etc. It has then been extended to also cover food product related aspects and nutritional indicators.

4. The Reconciled Database and Business Intelligence Services

The core of the data management component, as highlighted in Figure 3 is the reconciled database, which integrates different sources of information in order to serve BI services through a data warehouse.
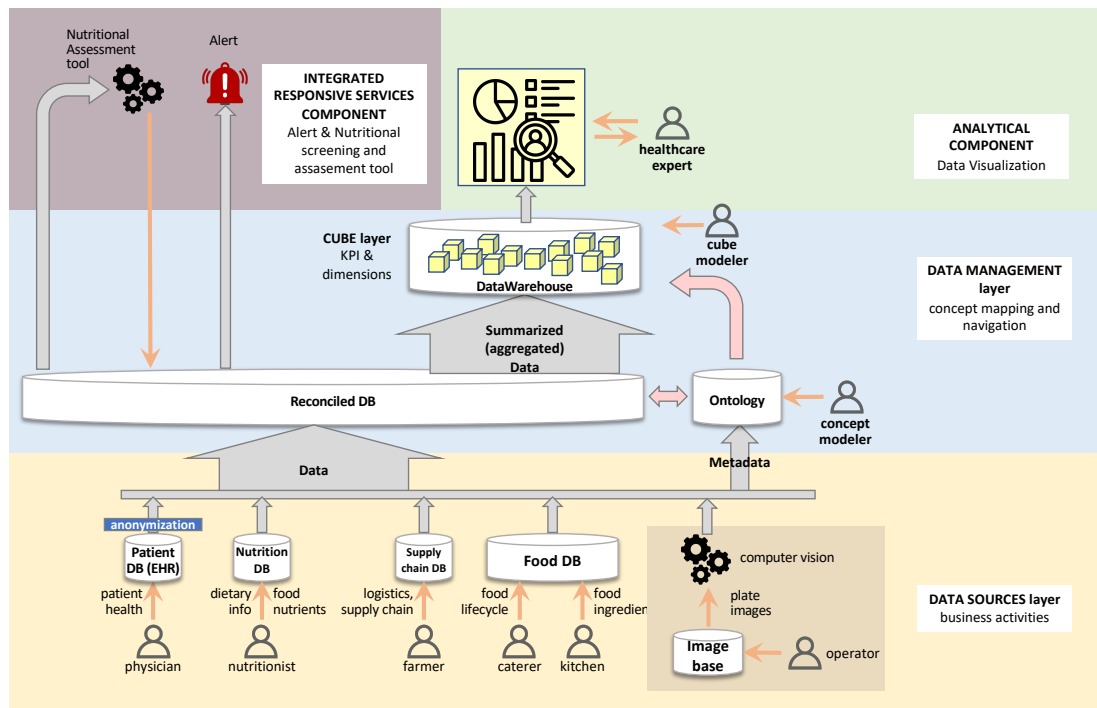
Figure 3: ADCATER solution Architecture. It is easy to note that the core of the data management component is the reconciled database, which integrates different sources of information in order to serve BI services through a data warehouse.

The reconciled database schema is created from the ontology via a semi-automatic process

based on approximation of the ontology language and transformation into digestible scripts by relational technologies. This process was supported by the OBDM tool Mastro [23].

Since this reconciled database is derived from the `ADCATER` ontology, it has several desirable properties. For example, it is centered on subjects, rather than applications, and thus is particularly suited for analytical purposes; it is rigorously documented; it is free from possible inconsistencies or other modeling issues; table and field names are related to the ontology terminology, so their semantics is clear and easily understandable by all stakeholders.

Regarding the population of the reconciled database the data are taken from several data sources such as the FoodFix application databases, those hospital repositories and nutritionists databases and archives. These data are then mapped against the tables in the reconciled database using standard techniques of data integration [24]. On top of the reconciled database, a BI layer is built based on Data Warehouse (DW) technologies. In a DW, the main players are the multidimensional objects, usually called cubes, which by aggregating data according to certain dimensions allow for analyses that would otherwise be difficult to implement using classical DBMs. These cubes, taking the terminology of the reconciled database are also aligned with that of the `ADCATER` ontology. The design of the cubes therefore was directly done by inspecting it.

5. Conclusion

In this paper, we have introduced the ontology `ADCATER`, which formalizing the medical and nutritional domain of the homonymous project, guides the implementation of the solution's data management module. The `ADCATER` ontology by harmonizing terminology and providing an inconsistency-free data layer enables the definition of a reconciled database to enable analysis using business intelligence technologies. As future work, it would be interesting to expand or integrate the knowledge with the help of other medical ontologies, such as those related to mental health issues, and consequently accommodate BI techniques to get a better view of patient status. From a data management and data quality perspective, it would be beneficial to integrate data preparation techniques to address issues like entity disambiguation [25, 26] and to extend data integration to include non-relational sources [27, 28], such as textual data [29, 30].

References

[1] N. Guarino, D. Oberle, S. Staab, What is an ontology?, Handbook on ontologies (2009) 1–17.

[2] A. Hogan, E. Blomqvist, M. Cochez, C. d'Amato, G. D. Melo, C. Gutierrez, S. Kirrane, J. E. L. Gayo, R. Navigli, S. Neumaier, et al., Knowledge graphs, ACM Computing Surveys (Csur) 54 (2021) 1–37.

[3] I. Horrocks, Ontologies and the semantic web, Communications of the ACM 51 (2008) 58–67.

[4] M. Lenzerini, Ontology-based data management, in: Proceedings of the 20th ACM international conference on Information and knowledge management, 2011, pp. 5–6.

[5] O. Corcho, F. Priyatna, D. Chaves-Fraga, Towards a new generation of ontology based data access, Semantic Web 11 (2020) 153–160.

[6] G. Xiao, D. Calvanese, R. Kontchakov, D. Lembo, A. Poggi, R. Rosati, M. Zakharyaschev, Ontology-based data access: A survey, International Joint Conferences on Artificial Intelligence, 2018.

[7] M. Console, M. Lenzerini, Data quality in ontology-based data access: The case of consistency, in: Proceedings of the AAAI Conference on Artificial Intelligence, volume 28, 2014.

[8] G. Xiao, L. Ding, B. Cogrel, D. Calvanese, Virtual knowledge graphs: An overview of systems and use cases, Data Intelligence 1 (2019) 201–223.

[9] D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, R. Rosati, et al., Ontology-based data access and integration, in: Encyclopedia of database systems, Springer, 2018, pp. 2590–2596.

[10] T. Cederholm, G. Jensen, M. Correia, M. C. Gonzalez, R. Fukushima, T. Higashiguchi, G. Baptista, R. Barazzoni, R. Blaauw, A. Coats, et al., Glim criteria for the diagnosis of malnutrition–a consensus report from the global clinical nutrition community, Journal of cachexia, sarcopenia and muscle 10 (2019) 207–217.

[11] H. M. Kruizenga, J. Seidell, H. C. de Vet, N. Wierdsma, et al., Development and validation of a hospital screening tool for malnutrition: the short nutritional assessment questionnaire (snaq©), Clinical Nutrition 24 (2005) 75–82.

[12] D. Lembo, V. Santarelli, D. F. Savo, G. D. Giacomo, Graphol: A graphical language for ontology modeling equivalent to OWL 2, Future Internet 14 (2022) 78. URL: https://doi.org/10.3390/fi14030078. doi:10.3390/FI14030078.

[13] P. Hitzler, A. Gangemi, K. Janowicz, Ontology engineering with ontology design patterns: foundations and applications, volume 25, IOS Press, 2016.

[14] K. A. Spackman, K. E. Campbell, R. A. Côté, Snomed rt: a reference terminology for health care., in: Proceedings of the AMIA annual fall symposium, American Medical Informatics Association, 1997, p. 640.

[15] M. Q. Stearns, C. Price, K. A. Spackman, A. Y. Wang, Snomed clinical terms: overview of the development process and project status., in: Proceedings of the AMIA Symposium, American Medical Informatics Association, 2001, p. 662.

[16] W. H. Organization, et al., International classification of diseases:[9th] ninth revision, basic tabulation list with alphabetic index, World Health Organization, 1978.

[17] A. Maercker, C. R. Brewin, R. A. Bryant, M. Cloitre, G. M. Reed, M. Van Ommeren, A. Humayun, L. M. Jones, A. Kagee, A. E. Llosa, et al., Proposals for mental disorders specifically associated with stress in the international classification of diseases-11, The Lancet 381 (2013) 1683–1685.

[18] S. d. Coronado, M. W. Haber, N. Sioutos, M. S. Tuttle, L. W. Wright, Nci thesaurus: using science-based terminology to integrate cancer research results, in: MEDINFO 2004, IOS Press, 2004, pp. 33–37.

[19] C. E. Lipscomb, Medical subject headings (mesh), Bulletin of the Medical Library Association 88 (2000) 265.

[20] T. A. Holton, V. Vijayakumar, N. Khaldi, Bioinformatics: Current perspectives and future directions for food and nutritional research facilitated by a food-wiki database, Trends in food science & technology 34 (2013) 5–17.

[21] C. Caracciolo, A. Stellato, A. Morshed, G. Johannsen, S. Rajbhandari, Y. Jaques, J. Keizer, The agrovoc linked dataset, Semantic Web 4 (2013) 341–348.

[22] D. M. Dooley, E. J. Griffiths, G. S. Gosal, P. L. Buttigieg, R. Hoehndorf, M. C. Lange, L. M. Schriml, F. S. Brinkman, W. W. Hsiao, Foodon: a harmonized food ontology to increase global food traceability, quality control and data integration, npj Science of Food 2 (2018) 23.

[23] L. Lepore, M. Namici, G. Ronconi, M. Ruzzi, V. Santarelli, The mastro ecosystem: Ontology-based data management from theory to practice, in: 2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE), IEEE, 2019, pp. 101–102.

[24] D. Calvanese, G. De Giacomo, D. Lembo, M. Lenzerini, A. Poggi, M. Rodriguez-Muro, R. Rosati, M. Ruzzi, D. F. Savo, The mastro system for ontology-based data access, Semantic Web 2 (2011) 43–53.

[25] R. Fagin, P. G. Kolaitis, D. Lembo, L. Popa, F. Scafoglieri, A framework for combining entity resolution and query answering in knowledge bases, in: Proceedings of the 20th International Conference on Principles of Knowledge Representation and Reasoning, KR 2023, Rhodes, Greece, September 2-8, 2023, 2023, pp. 229–239.

[26] R. Fagin, P. G. Kolaitis, D. Lembo, L. Popa, F. Scafoglieri, Combining entity resolution and query answering in ontologies: A formal conceptual framework (discussion paper), in: Proceedings of the 32th Italian Symposium on Advanced Database Systems, SEBD 2024, CEUR Workshop Proceedings, CEUR-WS.org, 2024.

[27] D. Lembo, F. M. Scafoglieri, Comparing state of the art rule-based tools for information extraction, in: Rules and Reasoning - 7th International Joint Conference, RuleML+RR 2023, Oslo, Norway, September 18-20, 2023, Proceedings, volume 14244 of *Lecture Notes in Computer Science*, Springer, 2023, pp. 157–165.

[28] D. Lembo, Y. Li, L. Popa, K. Qian, F. Scafoglieri, Ontology mediated information extraction with MASTRO SYSTEM-T, in: Proceedings of the ISWC 2020 Demos and Industry Tracks: From Novel Ideas to Industrial Practice co-located with 19th International Semantic Web Conference (ISWC 2020), Globally online, November 1-6, 2020 (UTC), volume 2721 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020, pp. 256–261.

[29] G. Ganino, D. Lembo, M. Mecella, F. Scafoglieri, Ontology population for open-source intelligence: A gate-based solution, Softw. Pract. Exp. 48 (2018) 2302–2330. URL: https://doi.org/10.1002/spe.2640. doi:10.1002/SPE.2640.

[30] D. Lembo, F. M. Scafoglieri, Ontology-based document spanning systems for information extraction, Int. J. Semantic Comput. 14 (2020) 3–26. URL: https://doi.org/10.1142/S1793351X20400012. doi:10.1142/S1793351X20400012.