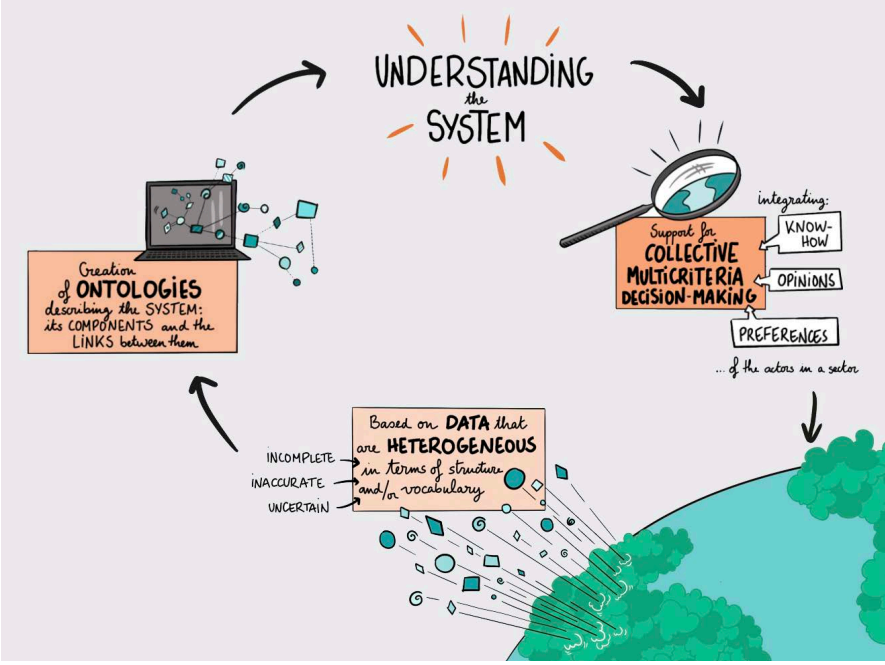
Synthèses

KNOWLEDGE ENGINEERING FOR DECISION SUPPORT IN THE FOOD AND BIOPRODUCT INDUSTRY

Integration of heterogeneous data through AI techniques

Patrice Buche, Pierre Bisquert, Rallou Thomopoulos, eds



éditions Quæ

Knowledge Engineering for Decision Support in the Food and Bioproduct Industry

Integration of Heterogeneous Data Through AI Techniques

Patrice Buche, Pierre Bisquert,
Rallou Thomopoulos, eds

Éditions Quæ

Quæ books in English

Remote Sensing and Spatial Modelling.
Applications to the Surveillance and Control of Mosquito-borne Diseases
Annelise Tran, Eric Daudé, Thibault Catry (Eds), 2025, 148 p.
https://doi.org/10.35690/978-2-7592-4102-6

Understanding Digital Agriculture.
Ten Years of Interdisciplinary Research at the #DigitAg Institute
Véronique Bellon-Maurel, Karine Gauche, Martha Lucia Enriquez, Nathalie Lyon-Caen (Eds), 2026, 202 p. https://doi.org/10.35690/978-2-7592-4360-0

To cite this work:
Buche P., Bisquert P., Thomopoulos R. (eds), 2026. *Knowledge Engineering for Decision Support in the Food and Bioproduct Industry. Integration of Heterogeneous Data Through AI Techniques*, Versailles, Éditions Quæ, 220 p. https://doi.org/10.35690/978-2-7592-4317-4

All book manuscripts are peer-reviewed by external reviewers.
This scientific evaluation process is described here:
https://www.quae.com/store/page/199/processus-devaluation.

The editorial process also relies on software for detecting similarities
and texts potentially generated by artificial intelligence.

Cover: Claire Hernandez, graphic facilitator.

© Éditions Quæ, 2026

Éditions Quæ
RD 10, 78026 Versailles Cedex, France
www.quae.com – www.quae-open.com

ISBN print: 978-2-7592-4316-7
ISBN ePub: 978-2-7592-4318-1
ISBN PDF: 978-2-7592-4317-4
ISSN: 1777-4624

The digital versions of this book are released under a CC-by-NC-ND 4.0 license
(https://creativecommons.org/licenses/by-nc-nd/4.0/).

Contents

Foreword 5
Johnny Beaugrand

Foreword 7
Nathalie Aussenac-Gilles

Introduction 11

Chapter 1. Domain Ontologies for Food and Biobased Products 15
Magalie Weber, Julien Cufi, Juliette Defossez, Stéphane Dervaux, Hervé Guillemin, Nicolas Guinet, Yannick Guyonvarch, Liliana Ibanescu, Emmanuele Piaud, Patrice Buche
1.1. Introduction 15
1.2 Proposed solutions 21
1.3 Outcomes 40
1.4 Conclusion and resources 47

Chapter 2. Task Ontologies for Product Processing and Characterization Processes 50
Liliana Ibanescu, Julien Couteaux, Julien Cufi, Stéphane Dervaux, Hervé Guillemin, Alrick Oudot, Emmanuele Piaud, Magalie Weber, Patrice Buche
2.1 Introduction 50
2.2 Proposed solutions 55
2.3 Outcomes 85
2.4 Conclusion and resources 88

Chapter 3. Methods for Ontology-Guided Extraction of Experimental Data in Scientific Articles 92
Martin Lentschat, Patrice Buche, Julien Cufi, Luc Menut, Mathieu Roche
3.1 Introduction 92
3.2 Proposed solutions 96
3.3 Outcomes 111
3.4 Conclusion and resources 117

Chapter 4. Reasoning About Process: Experimental Data and Technological Expert Know-how for Decision Support 121
Mélanie Münch, Cédric Baudrit, Julien Cufi, Julien Couteaux, Christophe Fernandez, Luc Menut, Alrick Oudot, Patrice Buche
4.1 Introduction 121
4.2 Proposed solutions 126
4.3 Outcomes 135
4.4 Conclusion and resources 144

Chapter 5. Methods for Collective Decision Support that Consider both Preferences and Contradictory Arguments from Agrifood Chain Stakeholders 148
Pierre Bisquert, Rallou Thomopoulos
5.1 Introduction 148
5.2 Proposed solutions 150
5.3 Outcomes 163
5.4 Conclusion and resources 168

Chapter 6. Designing Prospective Scenarios at the Scale of a Value Chain 171
Romy-Lynn Chaib, Catherine Macombe, Rallou Thomopoulos, Erwan Engel
6.1 Introduction 171
6.2 Proposed solution 172
6.3 Outcomes 176
6.4 Conclusion and resources 179

Chapter 7. Assessing the Impact of Participatory Interventions at the Individual Level 181
Reine Barbar, Rallou Thomopoulos
7.1 Introduction 181
7.2 Proposed solution 184
7.3 Outcomes 185
7.4 Conclusion and resources 185

Final Conclusion and Future Outlook 189

Appendix 1. Knowledge Engineering Models Used in the Book 192
Jérôme Fortin, Patrice Buche, Pierre Bisquert

Appendix 2. Legend for the Workflow Diagrams 200
Patrice Buche

References 202

Acknowledgments 216

List of authors 219

Foreword

In a world where data has become the lingua franca of science, business, and people, the ability to understand, link, and interpret information seems more critical than ever. Between the abundant quantitative data provided by sensors, experiments, and models from our labs and the wealth of qualitative knowledge produced by human expertise lies an imperfect interface that research is attempting to fix. It is exactly here that *Knowledge Engineering for Decision Support in the Food and Bioproduct Industry* comes in, striking a balance between rigorous measurements and nuanced meaning.

Based on an ambitious collective undertaking under the editorial coordination of Patrice Buche, Pierre Bisquert, and Rallou Thomopoulos, researchers at IATE (INRAE–Institut Agro–University of Montpellier), this book is first and foremost the story of a scientific and human adventure. It explores the many paths along which artificial intelligence, well beyond algorithmic performance, can become not only a tool for mediation between disciplines but also undoubtedly a lever of collective intelligence and hopefully a driver of responsible innovation in the agrifood and biobased sectors. This book offers a wide-ranging menu of options.

An integrative and open approach to knowledge

The first two chapters lay down the conceptual foundations of this multiauthor book. It presents the vision of open data integration, capable of linking experimental measurements, field observations, and expert knowledge. The authors show how semantic structures and ontologies – subjects on which they have been pioneers – provide the underlying architecture of meaning that guarantees interoperability and what could be called "the sustainability of knowledge."

Chapters 3 and 4 get to the heart of the methodological approach. They describe the techniques of artificial intelligence that are used, such as machine learning, symbolic reasoning, knowledge representations and natural language processing, and above all the way these tools are combined in a hybrid, explainable, and participatory approach, which enhances their potential. Readers will discover the finesse of data engineering that never separates the technique from the human. The fourth chapter in particular illustrates the power of the methodological solutions put forward through a series of tangible applications in food and biobased products. From characterizing biomass to formulating new materials, each case study demonstrates how combining the quantitative and qualitative creates new opportunities for decision-making, traceability, and innovation.

Chapters 5 and 6 explore a common shortcoming faced by many labs and researchers – the organizational and collaborative aspects of a project – and how bridges can be created between labs, companies, and policymakers. These chapters also look at how open platforms can be built where data, knowledge, and preferences can become a common good rather than simply a technical instrument. Reading between the lines,

we can discern the authors' conviction that to be successful, a scientific project depends as much on the quality of the algorithms as it does on teamwork.

Finally, the last chapter expands the conversation to the future. It examines the role of open and smart systems in the digital transformation of the bioeconomy sectors and the ethical responsibility of their designers. More than a conclusion, it is an invitation: an invitation to continue the dialogue between fields and stakeholders, to create tools that are ever more transparent, collaborative, and meaningful.

A book that encourages dialogue between science and society

This book goes much further than simply presenting results. It embodies a way of doing science differently, openly, thoughtfully, and firmly rooted in the reality of current transitions and the need for research that seeks solutions to the climate emergency we now find ourselves in. This book is for researchers, engineers, and students, but also business leaders and public policymakers trying to understand how artificial intelligence can drive the sustainable transformation of food and biobased systems.

Johnny Beaugrand
TRANSFORM Division Head
INRAE

Foreword

At a time when large neural network models seem to be the solution to all problems and are often considered the only effective contributions of artificial intelligence (AI), I am delighted and honored to preface this book, which perfectly illustrates the richness, diversity, and relevance of AI research far beyond the use or development of these models alone. In the approach put forward through the work of the ICO team at the IATE–INRAE–Institut Agro–University of Montpellier unit, human knowledge and practices are sources of inspiration, objects of transmission, and key elements in the completion of tasks in conjunction with digital resources, whether software, data, or knowledge graphs. The work reported in this book is an exemplary contribution to knowledge engineering, both because it provides insight into a team research process and because it presents remarkable and reusable issues and results. It also contributes to the agrifood domain with a large set of real case studies illustrating how decision-making can be supported for tasks like food processing or food packaging design.

The editors of this book, Patrice Buche, Pierre Bisquert, and Rallou Thomopoulos, are renowned researchers in the field of AI, particularly in knowledge representation and engineering and know-how modeling, mainly in the agrifood sector. Patrice Buche has spent his entire career in this field, first as a lecturer at AgroParisTech until 2002, and since then as a research engineer at the IATE unit. He specializes in ontology alignment and knowledge extraction from texts and tables with a view to constructing ontologies and knowledge graphs used in decision support systems. Pierre Bisquert joined the IATE unit as an INRAE researcher in 2014, bringing his skills in reasoning, logic, and argumentation, mainly existential logic, to develop approaches based on knowledge graphs and rules for data integration. He took over from Buche as head of the ICO team. Finally, Rallou Thomopoulos has been conducting AI research in the agrifood sector since obtaining her PhD from AgroParisTech. As researcher and then research director at INRAE within the IATE unit since 2004, she has focused on the representation and interrogation of fuzzy or imprecise knowledge, then on the representation of chains of arguments in the form of graphs within decision support systems.

The work presented in this book is the result of the efforts of the ICO team at the IATE unit, of which all three editors are members. It contributes to the team's two areas of research: first, the ontology-driven integration of heterogeneous data and knowledge to formalize the knowledge needed for decision-making, and second, argumentation and social choice to integrate the preferences of the actors concerned and manage their different points of view. I should emphasize that it is rare for a research group to be able to capitalize so coherently on the work of each of its members. This book is the result of a long-term and genuine team effort, with each chapter itself a collective achievement. It confirms the complementarity and consistency of the work carried

out over the years, namely the projects and theses conducted within the ICO team. The permanent members of the team, and especially the editors of this book, played a key role in orchestrating, giving meaning to, and choosing a relevant direction for these different contributions, and now in bringing them together in this work. We can also be proud that this English edition reflects the original organizational structure of the French research environment, where a team includes several researchers working on personal scientific projects in conjunction with a team goal. In a way, this book could even be considered an ambassador for this approach. Indeed, it shows us that this way of working, which differs from the more pyramidal and individualistic organization of the English-speaking world, creates specific dynamics, promotes stimulating exchanges, and leads to outstanding results.

The seven chapters that make up this book provide a coherent and incremental reading experience. The authors are current and former members of the ICO team and long-term collaborators. Each chapter describes a building block that illustrates how AI can support decision-making in the agrifood sector, from collecting data and factual knowledge to building models, creating support tools, and evaluating the impact of proposed solutions on individuals, teams, and value chains. The results of one chapter are later reused in more ambitious contributions in subsequent chapters. Reading is made easier by the consistent structure of each chapter, which details the research issues and proposed solutions before presenting the outcomes that can be reused by the scientific community.

Chapter 1 (Domain Ontologies for Food and Biobased Products) presents various research projects that integrate multidimensional data into knowledge graphs and align the vocabularies used in different heterogeneous data sources. These solutions rely on domain ontologies. The importance of the terminological aspect of these ontologies (labels, comments, and definitions) for better validation by domain experts and more efficient alignment with similar resources or domain documents is worth noting. The outcomes of this work are not only the ontologies themselves, but also the methods and tools used to manage them (e.g., the PO^2 vocabulary and data manager) and to explore multiple databases (MultiDB explorer), classify data, and align source content.

Chapter 2 (Task Ontologies for Product Processing and Characterization Processes) introduces task ontologies, which are ontologies dedicated to specific tasks. These ontologies focus on the processes studied by the ICO team, such as product processing (e.g., bread or flour production from wheat) and product quality assurance and maintenance. Again, the outcomes include the ontologies, processes, and tools used to design them, as well as how these ontologies are populated to represent both the processing processes described in documents and data tables and the operational know-how that is typically passed on orally in organizations.

Chapter 3 (Methods for Ontology-Guided Extraction of Experimental Data in Scientific Articles) describes methods for the ontology-guided extraction of experimental data from scientific articles, with this work illustrated using scientific papers about food packaging. These methods rely on natural language processing to extract semantic relations. The work makes an original contribution to the extraction of n-ary relations (relations with more than two arguments), which can be used to collect several properties of a target entity. The process involves two steps: identifying the arguments and combining them to represent an instance of an n-ary relation.

Chapter 4 (Reasoning About Process: Experimental Data and Technological Expert Know-how for Decision Support) takes things a step further. Given the ontologies, knowledge graphs, and related methods and tools, the objective is to compile information on how this knowledge is utilized in the reasoning process for decision-making purposes. The three proposed approaches rely on both data and human expertise. They use either tree-based decision-making models, models trained on databases, or flexible queries based on fuzzy sets to combine several expert models. This work is applied to recommendations to correct food defects, select food packaging, and support biocomposite packaging design processes.

In Chapter 5 (Methods for Collective Decision Support that Consider both Preferences and Contradictory Arguments from Agrifood Chain Stakeholders), the objective targets the decision-making process itself. The proposed solution is based on the representation of the process using arguments in accordance with argumentation theory and social choice, and the decision-making process itself is examined. Two tools are presented: DAMN (a methodology for representing formal arguments in a statement graph and computing a decision by traversing this graph) and MyChoice (a web-based application that supports collective decision-making by assessing the acceptability of different alternatives based on several explanatory arguments for these alternatives).

Chapter 6 (Designing Prospective Scenarios at the Scale of a Value Chain) broadens the scope even further, evaluating innovations (i.e., changes in the value chain) by assessing their value to society. Three types of scenario-based methods are proposed to deal with standard situations, adapt to remote contexts, and optimize the evaluation process. The approach considers the technical feasibility and acceptability to humans of the proposed decisions. One example of a project involves improving food safety for children, with an evaluation of the overall impact of the new food chain.

Finally, Chapter 7 (Assessing the Impact of Participatory Interventions at the Individual Level) shifts the focus from the global impact to the local impact on individuals, using personalized food as an example. This assessment is based on a workshop in which visually impaired participants are informed and encouraged to prepare their own food. This has a significant impact on their knowledge, self-efficacy, practical skills, and cooking habits.

In conclusion, I highly recommend reading this book for several reasons, both for the scientific quality of its original contributions and for the rigor of the scientific approach it presents. It demonstrates the need to combine symbolic and more automated learning-based approaches in AI to achieve scalable systems that truly serve their users and provide relevant assistance. It will also be of interest to readers in the agrifood sector, who will find avenues for further work and significant contributions to the transfer of know-how and the advancement of knowledge in this field.

Nathalie Aussenac-Gilles
CNRS Research Director
Head of the MELODI team at IRIT

Introduction

This book explores real examples of applications to propose "workflows," i.e., combinations of data and knowledge processing, to solve complex problems specific to the food and bioproduct industry. These workflows make it possible to collect, organize, share, connect, and regularly update different types of data and knowledge (e.g., expert opinions, experimental data, stakeholder preferences and opinions, digital models). The book tackles the methodological obstacle of reconciling knowledge and often imperfect data (of variable reliability, imprecise, incomplete or contradictory, and from heterogeneous sources) with learning, prediction, and decision-support methods and tools. Knowledge engineering, the main family of approaches used in this book, is the branch of artificial intelligence that deals with the acquisition, formal representation, and operationalization of expert knowledge for use in computer systems. It provides the methodologies and tools necessary to build knowledge-based systems by structuring domain expertise into formal models that support automated reasoning and decision-making (Pagels et al., 1984; Studer et al., 1998). This novel compilation of knowledge engineering methods and tools for the agrifood industry that allow semi-automatic reasoning on both quantitative data and qualitative knowledge constitutes the book's originality. It results from a selection of research conducted by the INRAE knowledge engineering team (known by its French acronym ICO) at the IATE Research Unit in Montpellier, France, and its network of collaborators, namely within CATI DIISCICO, a professional IT network that brings together engineers and scientists from INRAE. This fruitful collaboration led the team and its network to win the Open Science Research Data Award in 2023 for the category "Creating the right conditions for reuse" from the French Ministry of Research. The research presented in this book by computer scientists specialized in knowledge engineering is the result of a long-term collaboration with many agroresource processing experts belonging to the INRAE TRANSFORM department and associated partners. The authors of this book would like to take this opportunity to thank them for their support in helping us design the solutions proposed here. This research was also inspired by a long-term collaboration with cheesemaking value chains, and especially French Protected Designation of Origin (PDO) value chains. This collaboration led to the creation of the DOCAMEX association in 2023 at the annual assembly of the CNAOL, a national association for producers of geographically protected cheeses. Today, it brings together 20 PDO cheese value chains (70% of national production) and several ENIL dairy schools to ensure the maintenance of the co-produced DOCAMEX software tool by the ICO team with collaborators within CATI DIISCICO.

To cope with rapid changes in the agrifood sectors and the multifaceted challenges they face (environmental, societal, economic), stakeholders and decision-makers need a comprehensive and integrated view of the studied fields of application – the agrifood and bioproduct sectors – while being able to recommend actions at fairly detailed levels (e.g., optimal controlled parameter settings for biomass processing unit operations to achieve a given set of objectives, selection of innovative packaging

materials for a given food product, recommendation of technological actions to obtain a desired quality or to correct a defect). Decision-making must be based on available knowledge (data, expert statements, preferences and opinions of actors, numerical models) using a multicriteria approach, and it must consider the different and potentially conflicting points of view and preferences of the involved actors. This book presents a synthetic overview of workflows that have several specificities: (i) they are based on methods derived from artificial intelligence, (ii) they are implemented in innovative tools for capitalization and processing of data and knowledge, (iii) they are used to solve complex problems specific to the food and bioproduct industry, and (iv) their ultimate aim is to provide relevant collective decision-making support tools. These workflows are illustrated in several complex decision support applications (design of biocomposites for food packaging applications, food packaging selection for a given food, technological recommendations to preserve the quality of protected designation of origin cheeses, etc.).

This book focuses on the methods and tools used to implement these workflows. This presentation on two levels not only highlights the genericity of the proposed approaches but also ensures the work will remain relevant over time since it will not be dependent on a particular computer technology. Open resources are also provided (software sources, web/desktop applications, datasets, ontologies, thesauri, etc.) so readers can put the concepts presented in the book into practice and reuse/adapt them for other uses in the food and bioproduct industry or for other applications.

All the chapters are presented in the same way and broken down into four sections. Each chapter begins with a section that presents issues and challenges and identifies the research questions addressed in the chapter. Workflow solutions proposed to tackle those questions are compared to the state of the art. In Section 2, workflow solutions are detailed for each research question. Outcomes associated with the solutions are presented in Section 3. Section 4 concludes the chapter and lists associated resources available on a public repository (e.g., Recherche Data Gouv, INRAE GitLab, HAL, INRAE Nextcloud). For all chapters, 34 main scientific papers, 7 domain ontologies, 3 task ontologies, 12 data papers comprising 43 datasets, 10 other datasets, 16 software programs, 7 web/desktop applications, 15 videos, and 11 reports are available through open access.

The first part of this book, which is divided into four chapters, covers the integration of heterogeneous data sources guided by ontology and coupling with multicriteria decision support systems:

1. Domain Ontologies for Food and Biobased Products
2. Task Ontologies for Product Processing and Characterization Processes
3. Methods for Ontology-Guided Extraction of Experimental Data in Scientific Articles
4. Reasoning About Process: Experimental Data and Technological Expert Know-how for Decision Support

The second part, with three chapters, is dedicated to collective decision support:

5. Methods for Collective Decision Support that Consider both Preferences and Contradictory Arguments from Agrifood Chain Stakeholders
6. Designing Prospective Scenarios at the Scale of a Value Chain
7. Assessing the Impact of Participatory Interventions at the Individual Level

Interested readers will find a short introduction to logical foundations and standardized languages used in the chapters to implement proposed solutions in Appendix 1. The graphical representation of workflow solutions used in all chapters is presented in Appendix 2.

Chapter 1

Domain Ontologies for Food and Biobased Products

Magalie Weber, Julien Cufi, Juliette Defossez, Stéphane Dervaux, Hervé Guillemin, Nicolas Guinet, Yannick Guyonvarch, Liliana Ibanescu, Emmanuele Piaud, Patrice Buche

1.1. Introduction

1.1.1. Issues and challenges

Knowledge discovery in food and biobased products engineering is challenging because the research questions that must be addressed focus on both process performances and product uses. Examples of complex research questions include food or biobased product formulation, process optimization, and the environmental impact assessment of processes. Around the world, momentum is also building to bring production and consumption patterns together in a circular, more sustainable approach, known as the biobased economy or bioeconomy. According to the Organisation for Economic Co-operation and Development (OECD, 2021), the agrifood system is actually a multidimensional system that encompasses "all the elements and activities related to producing and consuming food, as well as their effects, including economic, health, and environmental outcomes." Biorefining is defined by the IEA Bioenergy Task 42 on Biorefineries (de Jong et al., 2012) as "the sustainable processing of biomass into a spectrum of marketable products (food, feed, materials, chemicals) and energy (fuels, power, heat)." The need to develop more integrative approaches to reconnect agriculture, environment, food, and health has been reaffirmed recently to cope with the challenges of agrifood system sustainability and the "OneHealth" approach.[1]

The science at play in connecting these different areas involves various disciplines, from agronomy to environmental sciences, socioeconomics, biochemistry, microbiology, nutrition, sensometrics, chemistry and physics, process engineering, and computing sciences. The data collected are by nature multidimensional, and large collections of experimental datasets, which come from various projects and with heterogeneous sources and formats, should be exploited. These data are also sometimes incomplete or imprecise due to the high uncertainty of measurements inherent in biological systems or analytical techniques. As a result, these data must be integrated, which involves giving them a unified structure and unified semantics before

1. https://www.inrae.fr/en/europe-and-world/one-health-people-animals-and-environment

modeling and reasoning. Moreover, additional external data are needed to enrich the datasets. This is the underlying motivation for the use of data matching and alignment techniques to identify vocabularies with similar descriptions in heterogeneous data sources. In this chapter, we look at two real examples to illustrate this objective. In the first example, the goal is to complete missing data (vitamin C, vitamin B12, and iron in food) from one source (CIQUAL, the French food composition database) using an external source (USDA FoodData Central database). In the second example, the aim is to track changes in food composition (e.g., sugar and salt content) found in heterogeneous sources by drawing from a common reference managed by OQALI (the French food observatory). The results obtained using these methods are given in Sections 1.3.2 and 1.3.3. In this chapter, classification and alignment are considered to be the same task in which two types of alignments are distinguished: equivalence or specialization.

We address two research questions in this chapter:

– Q1.1: How can the multiple dimensions of the agrifood system be integrated to design sustainable, healthy, and affordable food and biobased products?

– Q1.2: How can vocabularies defined in different data sources be matched for integration purposes?

1.1.2. State of the art and positioning, including presentation of available scientific papers describing our proposal

Ontologies are a powerful tool for integrating the many dimensions of food systems. An ontology is a formal representation of a set of concepts within a domain and the relationships between those concepts (Gruber, 1993; Uschold and Gruninger, 1996). Ontologies are used to model knowledge in a way that allows mutual understanding between experts and interoperability among knowledge-based systems and agents (not only humans but also machines). Ontologies enable the standardization of vocabulary and provide a framework for data integration, reasoning, and analysis. There is a broad spectrum of ontologies, ranging from top-level ontologies (foundational ontologies), intermediate-level ontologies defining basic models or generic standards (core ontologies or ontology design patterns), and application-level ontologies (task or domain ontologies). Top-level ontologies are fundamental models, which are the most formal abstractions. A core ontology defines the generic model that can be further specialized for different domains or fields of applications. A task ontology focuses on the actions or tasks that should be performed to achieve a goal, often independent from a specific domain. A domain ontology gathers concepts from a specific domain.

In this book, we focus on and distinguish different kinds of ontologies, namely core ontologies (intermediate level) and domain or task ontologies (application level). For example, the core ontology for product processing might comprise both concepts (e.g., transformation steps and components) and relationships (e.g., hasInput (resp. hasOutput)) that link an input (resp. output) component to a step in a generic way. A domain ontology focuses on the entities and relationships specific to a particular area of knowledge (the domain of interest). For instance, a food domain ontology might include concepts such as raw materials (e.g., wheat, milk), final products (cheesecake), nutrients (e.g., vitamin C), and unit processing operations (e.g., heating), and relationships between entities (e.g., durum wheat is a kind of wheat, or cheesecake is a kind of cake).

The core ontology and the task ontology can be intertwined, as the generic concepts of the core ontology can both define tasks and be specialized within a domain ontology for a given application. For example, the core ontology presented above was designed to represent processing tasks. Tasks are expressed in a generic way in the core ontology, and concepts from the domain ontology are more specific to the domain (e.g., wheat is a kind of component, and heating is a kind of processing step).

Research question Q1.1 is addressed by using a domain ontology, TransformON, which is described in Section 1.2.1. This ontology will cover the aspects related to the processing of food, feed, bioproducts, and biowaste, as well as the characterization of the properties of raw materials, primary derivatives, end products, and waste or residue from the treated biomass.

To build the domain ontology, we used the Linked Open Terms methodology (Poveda-Villalón et al., 2022). First, we considered the existing resources needed to cover the defined domain, and more specifically both food and nonfood applications of biomass processing, including food, feed, biobased products, and biowaste valorization into bioactive molecules or biobased compounds, matter, and energy.

Following a literature study, we found many existing classification and description systems about food (see Ireland and Møller, 2016, for a review). Food classification and food description systems have different goals. A classification system tends to group foods from broad to narrower terms with similar characteristics. "Milk," for example, can be classified under "dairy products," while "wine" would go under "alcoholic beverages." This system serves as a solution for facilitating information retrieval for end users. Meanwhile, a description system is a solution for the data originator/producer, who wants to describe the food as precisely as possible from different points of view, without the need to aggregate them. For example, the International Network of Food Data Systems (INFOODS) (Truswell et al., 1991) provides free-text food descriptions, a listing of features or entities that might influence food composition and that data providers are encouraged to record. However, it does not always include controlled terms to standardize the description and facilitate data retrieval. In this type of situation, thesauri can be used to ensure consistent indexing and retrieval (ISO, 2011, 2013). A thesaurus, in which vocabulary control is achieved by providing textual definitions and organizing hierarchical relationships, is a useful resource that can complement a coding system.

Here are some valuable existing resources for the food domain:

– *The AGROVOC*[2] *and the LanguaL™ thesauri* are well-known thesauri for the agrifood domain. AGROVOC is a controlled vocabulary covering all areas of interest to the Food and Agriculture Organization of the United Nations (FAO), including food, nutrition, agriculture, fisheries, forestry, and environment. AGROVOC was first developed in the early 1980s, and currently contains over 41,000 concepts organized in a faceted hierarchy; each concept may have labels in up to 42 languages. However, its facets and terms are not sufficient or specific enough to describe food products in the form in which they are consumed. LanguaL™ (Ireland and Møller, 2010) is an international thesaurus using faceted classification and whose origins stretch back to the 1975 US FDA CFSAN Factored Food Vocabulary. This vocabulary, known by its acronym FFV, was first used in an information retrieval system

2. https://www.fao.org/agrovoc/releases

called "SIREN," the Scientific Information and Retrieval Exchange Network, at the Center for Food Safety and Applied Nutrition (CFSAN) of the US Food and Drug Administration (FDA). The current version of LanguaL™ 2017 contains a total of 12,605 descriptors arranged in a hierarchical structure organized into 14 facets. In all, more than 40,000 foods from Europe, North America, and a number of countries are indexed in LanguaL™. For example, the USDA National Nutrient Database for Standard Reference (SR legacy) has been fully LanguaL™ indexed, whereas the CIQUAL French Food Composition Database[3] published by the French Agency for Food, Environmental and Occupational Health & Safety (ANSES) has included only the facet A to reference the product category for all foods and has only fully LanguaL™-indexed a subset of foods.

– *FoodEx2*, which was developed by the European Food Safety Authority (EFSA), is "a comprehensive food classification and description system aimed at covering the need to describe food in data collections across different food safety domains" (EFSA, 2015). The general aim was to provide the food details necessary to make exposure or intake assessments in a food safety context. FoodEx2 is an authoritative reference that complements the EFSA's Standard Sample Description for Food and Feed (EFSA, 2013) at the European level and is used by national agencies in European countries for database annotation and data exchange. A parent-child relationship exists between a core list food item and its related extended list food items. The terms in the core and extended lists may be aggregated in different ways according to the needs of the different food safety domains. Collections of additional terms (called "facet descriptors") are also available. The list of terms describes the group to which the food product belongs, as well as the product itself. A unique term code is linked to each list term or facet descriptor, so it can be described in an unambiguous way. The classification is mostly based on the treatment used to obtain new food product categories and considers three types of food categories: raw materials, derivatives/ingredients, and composite foods. Raw materials are directly obtained from living sources after harvesting or slaughtering. Derivatives/ingredients are obtained by applying a process to raw materials. Composite foods are foods made with several raw materials or derivatives/ingredients. FoodEx2 also provides a list of feed for different groups of animals, which was also in our scope.

– *FoodOn*[4] is an ontology built to be interoperable with the international Open Biological and Biomedical Ontology Foundry (OBO Foundry[5]) and to represent entities that play a "food role" (Dooley et al., 2018). When it launched in 2018, FoodOn was largely based on LanguaL™ categories that were further converted in the ontology environment. The process of creating logical definitions started with a large number of food products that were introduced by way of the LanguaL™-indexed SIREN database. FoodOn provides a controlled vocabulary that can be used by both people and computers to name all parts of animals, plants, and fungi able to bear a food role for humans and domesticated animals, as well as derived food products and the processes used to make them. The aim is to develop semantics for food safety, food security, the agricultural and animal husbandry practices linked to food production, and culinary, nutritional, and chemical ingredients and processes.

3. https://ciqual.anses.fr/
4. https://FoodOn.org/
5. https://obofoundry.org/