Ontology based smart search within the existing consumer data platform

Deliverable in KB-38 Data Analytics for Food and Consumer-related Research

Robbert Robbemond

WAGENINGEN
UNIVERSITY & RESEARCH

Our team

KB-38 Data Driven and High-Tech Innovations

Project 1: Data Analytics for Food and Consumer-related Research

Work package 2: Linked Consumer Behaviour Data Services (WEcR)

Ireen Raaijmakers

Domain expert Consumer Research

Jeroen Teurlings

Software Developer

John Doornbos

Software Developer

David Jagtman

Data Engineer/Scientist

Robbert Robbemond

Product & Project Manager

Contents

Back-ground

Ontology based smart search

Enhancing Consumer Behavior Research through establishing a domain ontology

Harmonization in data accessibility and interoperability

The primary objective of this ontology in consumer research is to foster harmonization in data accessibility and interoperability. By establishing a standardized framework for organizing and categorizing data within the domain, this ontology aims to enhance collaboration, streamline data access, and facilitate seamless interaction between various datasets and systems.

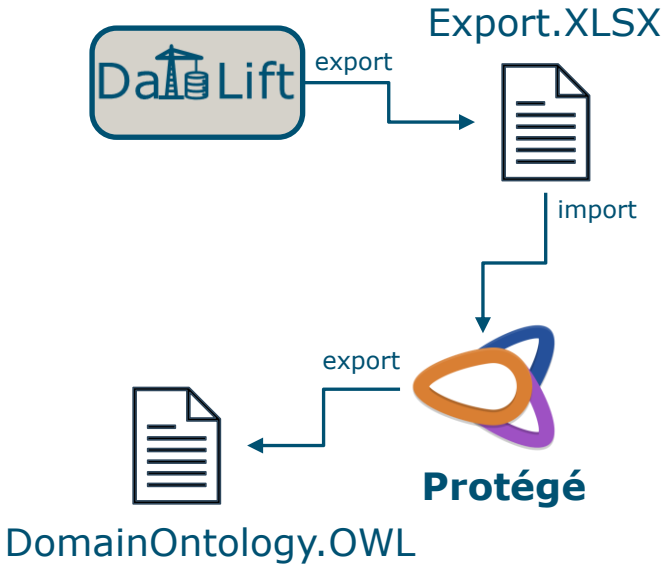
Understanding consumer behavior

Through a unified structure, it aspires to break down data silos and bridge the gap between disparate sources of information. This harmonized approach not only simplifies data retrieval but also promotes compatibility across different tools and platforms utilized in consumer research. Ultimately, the goal is to enable researchers, analysts, and stakeholders to efficiently access, interpret, and leverage data insights, fostering a more unified and comprehensive understanding of consumer behavior.

Back-ground

Ontology based smart search

Enhancing Consumer Behavior Research through establishing a domain ontology

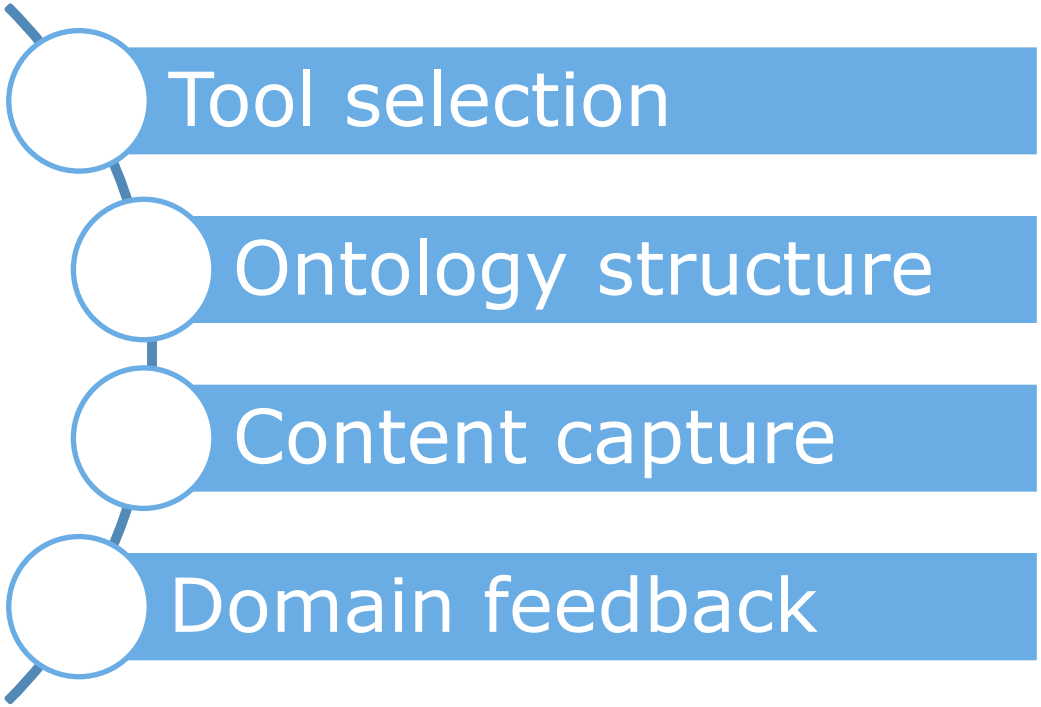


Pioneering a Domain Ontology for Consumer Research

This deliverable provides a pioneering effort in consumer research, developing an ontology tailored to the content within Datalift. Beyond structuring data, this ontology lays the groundwork for a domain-specific framework, potentially revolutionizing consumer research methodologies. By organizing Datalift's content, it aims to unlock profound insights, offering a structured approach to analyze consumer behavior and potentially reshape strategies in the dynamic consumer market.

Approach

Steps for Ontology Development



WAGENINGEN
UNIVERSITY & RESEARCH

Results

Tool Selection

Protégé selected as ontology editor

Tool / Characteristic:	User-Friendly Interface	Compatibility with Industry Standards	Extensibility and Customization	Visualization and Modeling Capabilities	Tool Adoption Level	License Type
Protege	4	5	4	4	5	Freeware (Open-source)
TopBraid Composer	4	4	4	4	3	Commercial
PoolParty	4	4	4	4	3	Commercial
Ontotext	4	4	4	4	4	Commercial

- **User-Friendly Interface:** All tools exhibit a high level of user-friendliness, simplifying ontology creation and management with intuitive interfaces.
- **Compatibility with Industry Standards:** Protege stands out with robust support for various ontology languages, while other tools maintain strong compliance with prevailing standards.
- **Extensibility and Customization:** All tools allow customization, catering to specific project needs through plugins or integrations.
- **Visualization and Modeling Capabilities:** Each tool provides advanced features for visualizing complex ontological structures effectively.
- **Tool Adoption Level:** Protege, an open-source tool, enjoys the highest adoption rate within the ontology development community. Other tools have moderate adoption levels.
- **License Type:** Protege operates as freeware (open-source), while TopBraid Composer, PoolParty, and Ontotext are commercial tools.

Results

Ontology structure

Upper ontology approach selected for structuring ontology

Top Level Ontology

Common Vocabulary	A top-level ontology provides a standardized set of foundational terms and concepts that can serve as a common vocabulary for different ontologies. This reduces ambiguity and ensures that entities are represented consistently across various knowledge domains.
Consistency	By adhering to a top-level ontology, ontologies developed in different domains can maintain uniformity in terms of how they represent concepts, properties, and relationships. This consistency enhances the quality and reliability of data integration.
Interoperability	Using a top-level ontology promotes seamless integration between disparate ontologies. It facilitates data exchange, sharing, and communication among systems that might have different underlying structures.
Semantic Alignment	A top-level ontology helps ontologies align their semantics, meaning that concepts and relationships are understood in the same way across different domains. This alignment improves the accuracy of data integration and reasoning.
Simplifies Mapping	When ontologies need to be combined or integrated, a top-level ontology provides a reference point for mapping concepts and relationships between them. This simplifies the process of aligning different ontologies.
Efficient Integration	Integrating ontologies that adhere to a common top-level ontology is generally more efficient than trying to merge ontologies with divergent structures. This saves time and effort in data harmonization.
Cross-Domain Understanding	Top-level ontologies offer a broader perspective that facilitates understanding across different domains. Users from various fields can collaborate more effectively when they share a foundational understanding of concepts.
Reduces Redundancy	A top-level ontology can define common concepts and relationships, which reduces the need for domain-specific ontologies to duplicate these definitions. This streamlines ontology development and maintenance.
Semantic Enrichment	A top-level ontology can enrich domain-specific ontologies by providing higher-level contextual information. This enhances the semantic richness of ontologies and improves their overall usefulness.
Interdisciplinary Research	When different fields use a common top-level ontology, interdisciplinary research becomes more accessible and effective. Researchers can draw connections between domains and leverage shared concepts.
Scalability	As the number of ontologies grows, maintaining consistency and interoperability becomes more challenging. A top-level ontology helps manage this complexity by offering a stable reference framework.
Alignment with Standards	Many standards and best practices recommend or require the use of a top-level ontology to ensure semantic interoperability. Adhering to such standards enhances the compatibility of ontologies within broader ecosystems.
Future-Proofing	A top-level ontology can provide a foundation that accommodates future changes and additions. It offers stability while allowing flexibility for extending ontologies to address new requirements.
Easier Communication	When domain experts communicate with each other or with stakeholders outside their field, a top-level ontology helps ensure that concepts are understood consistently, regardless of the audience's expertise.
Reusability	Concepts and relationships defined in a top-level ontology can be reused across different ontologies, promoting modularity and reducing redundancy.

Results

Ontology structure

BFO selected as most suitable top-level ontology

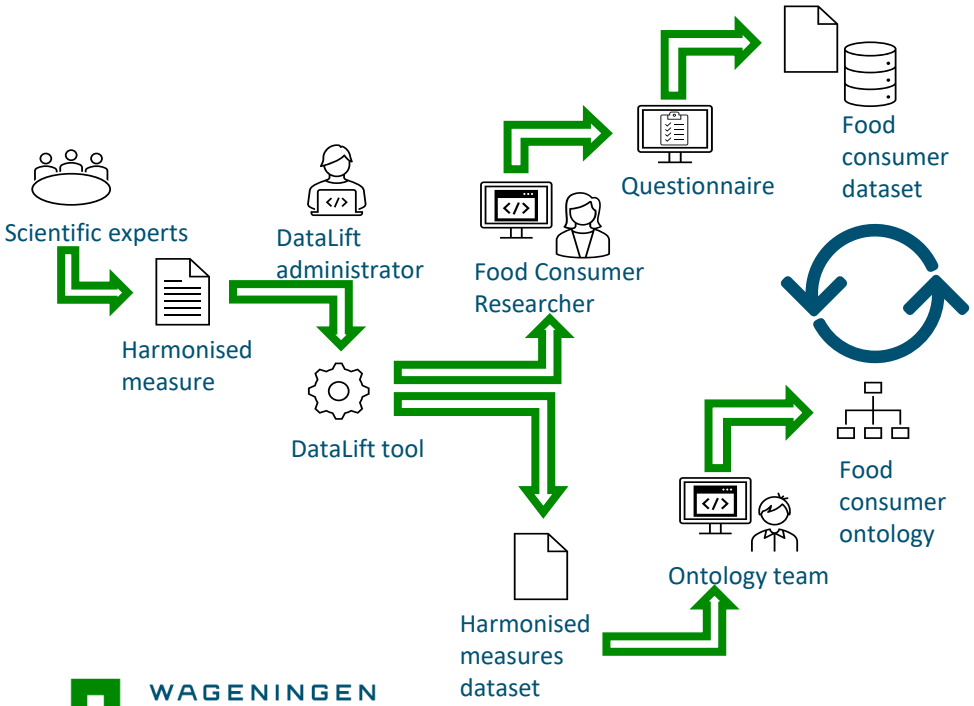
Characteristic:	Formalism	Granularity	Applicability	Community Support	Adoption as Foundation for Food, Nutrition, Biomedical, and Life Sciences
Basic Formal Ontology (BFO)	4	3	5	5	4
Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE)	4	4	4	3	3
General Formal Ontology (GFO)	3	4	4	4	2
Unified Foundational Ontology (UFO)	5	5	4	3	3

- **Formalism:** Assesses the structural complexity and formal nature of the ontology. Higher scores indicate a more intricate or versatile structure.
- **Granularity:** Indicates the level of detail and specificity the ontology provides. Higher scores imply a more detailed representation suitable for specialized domains.
- **Applicability:** Reflects the ontology's versatility and usefulness across various domains. Higher scores suggest broader applicability.
- **Community Support:** Indicates the level of acceptance and adoption within the ontology community. Higher scores suggest greater community endorsement and usage.
- **Adoption as Foundation for Food, Nutrition, Biomedical, and Life Sciences:** Evaluates how well each upper-level ontology can serve as a structural foundation specifically for domain-specific ontologies in food, nutrition, biomedical, and life sciences which are highly related to our consumer research domain. Higher scores signify better adaptability and applicability for these specialized domains.

Results

Content capture

Method designed for content capture



Standardized Measures through DataLift
Our ontology development process relies on DataLift, the Consumer Data Platform, to facilitate standardized measures. These measures, encapsulated in an XLSX file, define questionnaire elements uniformly for consumer researchers.

Ontology Integration with BFO Structure
Our ontology team accesses this file, integrating its contents into our ontology structured on BFO. To automate this process, a custom script will extract and input the XLSX data directly into the ontology.

Facilitating Interconnected Data Structures
This streamlined approach ensures a seamless link between community measures and our ontology, fostering interconnected and standardized data structures. It aligns community and ontology team efforts for consistency and collaboration within the domain.

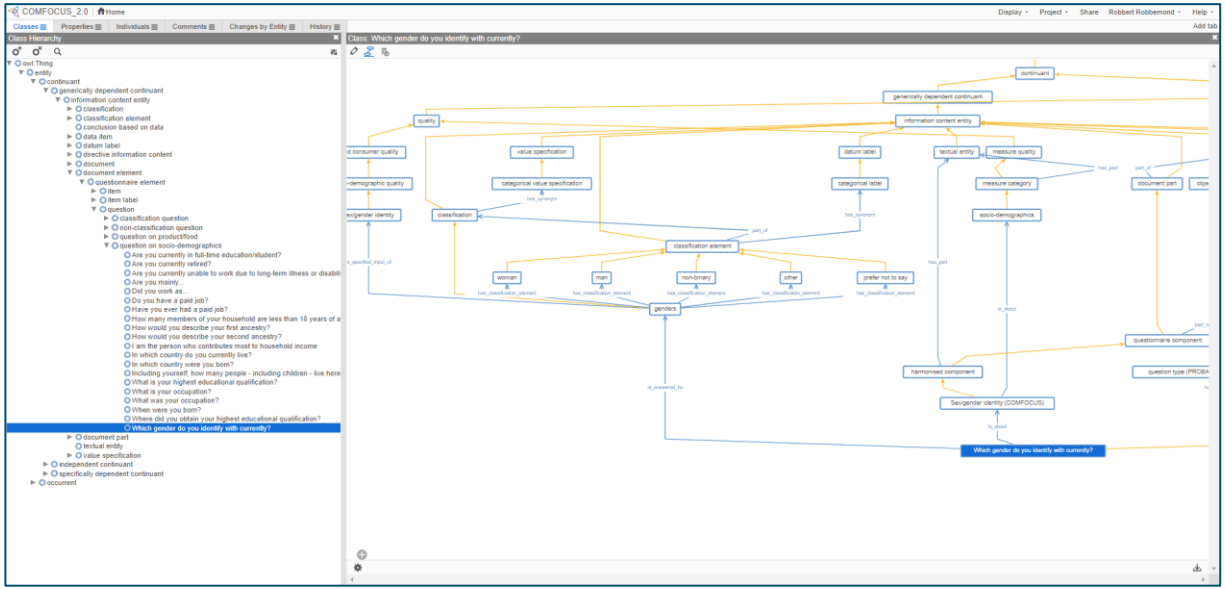
WAGENINGEN UNIVERSITY & RESEARCH

Results

Domain feedback

Examples created for script development

- **Examples created for script development**
- **Examples presented to a selection of domain experts**
- **Final version will be scrutinized by ontology expert (2024)**

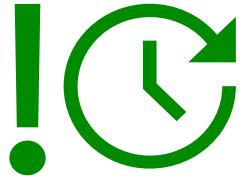


Conclusions

What output?	What outcome?	What impact?
Protégé selected as ontology editor	Widespread support from the ontology development community	By establishing a standardized framework for organizing and categorizing data within the domain, this ontology aims to enhance collaboration, streamline data access, and facilitate seamless interaction between various datasets and systems.
Upper ontology approach selected for structuring ontology	Consistent and meaningful communication between different ontologies and knowledge systems	
BFO selected as most suitable top-level ontology	Alignment with food, nutrition, biomedical, and life sciences	
Method designed for content capture	User-friendly mechanism available to synch domain standards with ontology	
Examples created for script development	Team aligned by shared vision on ontology framework	

Next steps

Future developments



What output?	What outcome?	What impact?
Ontology updated with content gathered by script	Ontology reflects domain standards and metadata	By establishing a standardized framework for organizing and categorizing data within the domain, this ontology aims to enhance collaboration, streamline data access, and facilitate seamless interaction between various datasets and systems.
Thorough domain evaluation feedback gathered and processed	Domain community involved in ontology development	
Ontology publication	Ontology available for entire community and beyond	

Thank you for your attention!