Holograph: a generic RDF schema to handle data from agroecological holobionts

Marie Lahaye[1,*,†], Alice Mataigne[2,†], Edmond Berne[3,4,5,†], Matéo Boudet[1,2], Olivier Dameron[2], Valentin Loux[3,4,5], Anne-Françoise Adam-Blondon[5,6], Olivier Rué[3,4] and Fabrice Legeai[1,2,*]

[1]*UMR 1349 IGEPP, INRAE, Le Rheu 35650, France*
[2]*Univ Rennes, Inria, CNRS, IRISA - UMR 6074, F-35000 Rennes, France.*
[3]*Université Paris-Saclay, INRAE, MaIAGE, 78350 Jouy-en-Josas, France*
[4]*Université Paris-Saclay, INRAE, BioinfOmics, MIGALE bioinformatics facility, 78350 Jouy-en-Josas, France*
[5]*IFB-Core, Institut Français de Bioinformatique (IFB), CNRS, INSERM, INRAE, CEA, 94800 Villejuif, France*
[6]*URGI, INRAE, Université Paris-Saclay, 78026 Versailles, France*

Keywords
Data integration, ontologies, holobionts, RDF, agroecology, metagenomics, environmental data

1. Introduction

Holobiont aims to represent an host - an animal or a plant - and its microbiota as a whole. Agroecological projects studying holobionts involve identifying and characterizing the communities of living organisms associated with cultivated plants or livestock. The composition of the microbiota is generally obtained using metabarcoding methods, resulting in abundance tables of Amplicon Sequence Variants (ASVs), markers of the observed taxa. In addition, in plant holobiont studies, many variables are collected in the field or in laboratory to describe precisely the plant phenotypes (yield, diseases), its environment (presence of bioagressors and weeds, agricultural practices, soil biochemistry) and climatic data (daily temperatures or relative humidity).

To enable the scientific exploitation of this data related to various hosts, their microbiota and their environment, it is essential to formally modelize it as a knowledge graph with domain specific ontologies. In the frame of the french program MUDIS4LS-IS5 (Mutualised Digital Spaces for FAIR data in Life and Health Science), we developped, Holograph, a generic RDF-schema partly based on the SOSA and I-ADOPT core-ontologies. this schema was applied to store data from a large project.

2. The Holograph schema

The Holograph schema is subdivised in 4 parts describing: (i) the microbiota (ii) the sample settings (iii) the observation (iv) and the variables.

The microbiota part represents samples, sequencing runs and bioinformatics analysis with their associated metadata, as well as results: ASV abundances and taxonomic assignations. These assignations are stored as URIs linking to the NCBI Organismal Classification (NCBITAXON) allowing to query the dataset at any taxonomic level (e.g. species, genus), through federated queries.

The "sample settings" part represents the contextual environment of the metagenomic samples. It describes the biological compartment from where the sample has been extracted, the corresponding host (e.g. plant) and geographical information (e.g. plot where the plant is cultivated and the field containing the plot). The location nesting is managed by the GeoSPARQL [1] ontology, and location

---

SWAT4HCLS: Semantic Web Applications and Tools for Health Care and Life Sciences, March 23–26, 2026, Amsterdam, The Netherlands

[*] Corresponding author.

[†] These authors contributed equally.

marie.lahaye@inrae.fr (M. Lahaye); fabrice.legeai@inrae.fr (F. Legeai)

types are described using some of the following ontologies: ENVO (Environment Ontology), AGRO (Agronomy Ontology), PO (Plant Ontology) or AGROVOC [2].

Measurements (e.g. phenotypic, environmental, agronomical and climatic descriptors) made on features of interest (in this case the various elements composing the sample settings, such as host and location), are modelized in the observation part using the Semantic Sensor Network Ontology (SOSA) [3].

Variables linked to observations in SOSA are described using the I-ADOPT Framework ontology [4] which break them down into elementary elements: a property (a characteristic of the object of interest), an entity (an object of interest) and a constraint (which limits the scope of the observation).

3. Application

The model was implemented on the data of the DeepImpact (https://anr.fr/ProjetIA-20-PCPA-0004) project which aims to understand the impact of microbial diversity on the soil, as well as physico-chemical and environmental conditions on health and yield of wheat and rapeseed crops. This dataset includes hundreds of agro-environmental descriptors (e.g. bioagressors inventories, plant phenotypic traits, ...), climatic data and metabarcoding data (ASV abundances and taxonomic assignation) from leaves, roots, rhizosphere and bulk soil samples.

The entirety of the DeepImpact (https://anr.fr/ProjetIA-20-PCPA-0004) dataset was integrated in AskOmics (https://askomics.org/), a RDF database with a web interface which allows the user to query the database by creating a path following the abstraction of the model. To do so, we created CSV templates to describe the differents objects and observations, and developed a tool to integrate these data - as well as the BIOM files in which metagenomics data are described - in AskOmics.

4. Conclusion/Perspectives

In conclusion, we developed a generic model to describe plant holobionts. It was successfully implemented on the DeepImpact project. While it is currently tested on livestock holobionts usecases, this model can be shared to anyone wanting to structure its own holobiont's data.

To further facilitate data exploration and extraction, a new user-friendly web interface is currently under development and will allow users to query the data without requiring knowledge of the underlying model abstraction.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] N. J. Car, T. Homburg, Geosparql 1.1: Motivations, details and applications of the decadal update to the most important geospatial lod standard, ISPRS International Journal of Geo-Information 11 (2022). URL: https://www.mdpi.com/2220-9964/11/2/117. doi:`10.3390/ijgi11020117`.

[2] I. Subirats-Coll, K. Kolshus, A. Turbati, A. Stellato, E. Mietzsch, D. Martini, M. Zeng, Agrovoc: The linked data concept hub for food and agriculture, Computers and Electronics in Agriculture 196 (2022) 105965. URL: https://www.sciencedirect.com/science/article/pii/S0168169920331707. doi:`https://doi.org/10.1016/j.compag.2020.105965`.

[3] K. Janowicz, A. Haller, S. J. Cox, D. Le Phuoc, M. Lefrançois, Sosa: A lightweight ontology for sensors, observations, samples, and actuators, Journal of Web Semantics 56 (2019) 1–10. URL: https://www.sciencedirect.com/science/article/pii/S1570826818300295. doi:`https://doi.org/10.1016/j.websem.2018.06.003`.

[4] B. Magagna, G. Moncoiffé, A. Devaraju, M. Stoica, S. Schindler, A. Pamment, R. I.-A. WG, Interoperable descriptions of observable property terminologies (i-adopt) wg outputs and recommendations, 2022. URL: https://doi.org/10.15497/RDA00071. doi:10.15497/RDA00071.