European Plant Phenotyping Network

EPPN 2020

D5.4b Two Training Schools (linked to workshop with technology developers)

Online training on object identification and variables naming

Isabelle Alic, Jean-Eudes Hollebecq

ject has received funding from the European Union's Horizon 2020 research and innovation programme under grant a ent No 731013. This publication reflects only the view of the author, and the European Commission cannot be held r ble for any use which may be made of the information contained therein.

Document information

EU Project N°	731013	**Acronym**	EPPN2020
Full Title	European Plant Phenotyping Network 2020		
Project website	www.eppn2020.plant-phenotyping.eu		

Deliverable	**N°**	D5.4b	**Title**	Two Training schools (linked to workshop with technology developers)
Work Package	**N°**	5	**Title**	Networking

Date of delivery	**Contractual**		31/01/2020 (Month 33)	**Actual**	18/10/2021 (Month 54)
Dissemination level	X	**PU Public, fully open, e.g. web**			
		CO Confidential, restricted under conditions set out in Model Grant Agreement			
		CI Classified, information as referred to in Commission Decision 2001/844/EC.			

Authors (Partner)	INRAE			
Responsible Author	**Name**	Jean-Eudes Hollebecq	**Email**	jean-eudes.hollebecq@inrae.fr

Version log			
Issue Date	**Revision N°**	**Author**	**Change**
02/09/2021	1	Jean-Eudes Hollebecq	First draft
15/09/2021	2	François Tardieu	Validated version
19/10/2021	3	Isabelle Alic	Some added corrections. Final Version

Executive Summary

This deliverable describes the second training sessions with technology developers, operated as part of JRA3. This course was organized in two online sessions on 16th and 20th of April 2021.

The objective was to propose a tool to generate identifiers for objects and names for variables in order to improve the interoperability of these concepts. A second objective was to allow discussion about the issues raised by real cases for variables naming

We targeted two persons per local infrastructure, namely the person in charge of data management in the corresponding infrastructure and the responsible of the infrastructure. These infrastructures were those of EPPN2020 but also those of the EMPHASIS support group. Specialists of data management in SMEs also attended.

The training was using an on-line tool for generation of URIs and for variable naming, developed during the project, in order to facilitate the discussions and the questions.

The first of the two sessions was centred on URI generation and the model for variable naming. At the end of the first session. Some homework have been given in order to prepare the next session.
The second session was a practical session were attendee could practise and validate their homework, real variables used by the attendees were introduced in the system. Multiple questions were answered by the organizing team as well as the audience itself, making it lively and self-sufficient.

Main Results:

Two sessions took place, 2h each, the 16th and 20th of April, with some time in-between to assimilate, think about it, and provide the second session with concrete examples to discuss. The training was held online and gathered 39 participants over the 2 sessions.

Authors/Teams involved:

- Isabelle Alic (INRAe)
- Jean-Eudes Hollebecq (INRAe)
- Peter Roos (WUR)
- François Tardieu (INRAe)

Table of content

1. MOTIVATION

Tracking all objects involved in a phenotyping experiment (e.g. plants, pots, sensors) and representing relationships between them is essential in a high throughput context where thousands of plots, plants or sensors are involved. This requires a proper strategy allowing to identify individually each specific object as well as semantic properties for creating relationships between such objects.

The URI schema was adopted within the consortium as the schema for identification. URIs follow a standard used in genetics, chemistry, Internet of Things, life sciences, etc. They identify objects (virtual or real) with a syntax that allows non-ambiguousness, unicity, persistence, stability and resolvability.

Variables need to be identified as well in order to make them interoperable. It was found that a dictionary or an ontology of all the existing variables is a goal impossible to achieve. Instead, a pattern was proposed to create new variables, compatible with MIAPPE where variables are detailed in “Trait-Method-Unit”. The MIAPPE model was improved, by splitting “Trait" into “Entity” and “Characteristic”. These results in a four components model: Entity-Characteristic-Method-Unit.

A TRAINING ON SEMANTIC WEB APPLIED TO PHENOTYPING

Identification and stable data models are important for findability and accessibility. In order to achieve the next steps of FAIR data we need to enhance interoperability. This can be achieved by linking data to existing concepts to give more information. Concepts are described in ontologies and we have selected some ontologies where the descriptions and metadata.

For the variables especially, the model is designed to help understanding of variables by the whole community. Interoperability and reusability cannot be achieved if the different parts of the model do not have a proper definition. In order to make this model meaningful, we have to make links between the components and the different ontologies that already exist to describe the entities, the physical quantities and qualities measured, the unit involved in such measurements. To do so we imported the concepts from the Ontology of Units of Measurements (OM).

2. TRAINING SESSIONS

The first session (16th April) was dedicated to :
- Theory details for URI generation, explanation of the URI schema and how to generate URI for different types of objects. Presentations were given, and followed by discussions on identifiers. The latter are created using an incremental schema, which ease the work for everyday manipulation. The model can also be extended to images or individual data points, generated in millions or even more. The pros and cons of different approaches were discussed

- Variables names already have a schema declared in MIAPPE. However, this model still is imperfect because desired traits are most often absent of the existing ontologies. For instance, "leaf height" or "meristem temperature" are not reported in ontologies, but "leaf" or "meristem" on one hand, "height" or "temperature" on the other hand can easily be found in ontologies. Hence, it was discussed that "Trait-Method-Unit" can be split into "Entity-Characteristic-Method-Unit". Discussions followed about combining the same entity with different characteristics and physical quantities. The ontologies used to describe the characteristics and the units were presented, in particular the Ontology of Units of Measurements (OM).

At the end of the first session, homework were given to attendees. They have been asked to provide some variables from real datasets and transform them into the new data model. They have been asked to insert a list of objects, like plants, plots, leaves etc. The tool is accessible under the following address: http://138.102.159.36:8082/app/ .

The second session (20th April) was a hands-on session. Data from the attendees were inserted and URI generated. The tool proposed proved to facilitate the adoption of an information system, starting with the identification of objects.

Questions raised for data validation and curation. As the tool proposed was open for the whole community, some declarations, especially for variables will need some curation. The way URI are declared makes it possible to validate the concepts based on the user who created it.

The training ended with a short demonstration of how the tools tested by trainees are inserted in the information system PHIS.

3. EVALUATION BY PARTICIPANTS

A short discussion was held at the end of the training session about what was achieved, followed by a formal questionnaire. The main points are:

- Organisation and content of the training:
  - Feedback on the organisation (duration of sessions, communication, facilitation of sessions) was positive
  - The main objectives of the participants, which were achieved, were to:
    - Understand variables concepts
    - Understand URI concepts
    - Discover a tool to generate variables
    - Discover a tool to generate URIs
    - Exchange with other people from the community about the challenges of variables and URIs

- Information systemSome participants would have liked more time to be dedicated either to the information systemPHIS, or to the presentation of ontologies and semantic resources

PARTICIPANT LIST

Name	**mail**
Giacomo mele	giacomo.mele@cnr.it
Tatu polvinen	tatu.polvinen@helsinki.fi
Hanna cwiek	hcwi@igr.poznan.pl
Stephan summerer	stephan.summerer@alsia.it
Juan herrera	juan.herrera@agroscope.admin.ch
Kristiina himanen	kristiina.himanen@helsinki.fi
George sainsbury	george.sainsbury@adelaide.edu.au
Tomasz Mroz	
Merlijn morise	merlijn.morisse@psb.vib-ugent.be
Sanja Mikic	sanja.mikic@ifvcns.ns.ac.rs
Nicolas Blanchet	nicolas.blanchet@inrae.fr
Sahameh Shafiee	sahameh.shafiee@nmbu.no
Felicia Maviane Macia	felicia.maviane-macia@inrae.fr
Sylvain Prigent	sylvain.prigent@inrae.fr
Jana Barbro Winkler	bwinkler@helmholtz-muenchen.de
Sigalas, Alexandros	alexandros.sigalas@helmholtz-muenchen.de
Denis Vile	denis.vile@inrae.fr
Kerstin Neumann	neumannk@ipk-gatersleben.de
Mickael Lamboeuf	mickael.lamboeuf@inrae.fr
David Benaben	david.benaben@u-bordeaux.fr
Peter Roos	peter.roos@wur.nl
Hamza Ikiou	
Maxilmilian Hart	
Inês Chaves	ichaves@itqb.unl.pt
Farzaneh Kazemi	
Frédéric Salvi	<f.salvi@terresinovia.fr>
Benoit Jaillais	Benoit.Jaillais@inrae.fr
Bendy Latortue	
Ethan Stewart	
Nuria De Diego	
P. Vermeulen	p.vermeulen@cra.wallonie.be
Jakub Jez	
François Tardieu	
Sahameh Shafiee	
Sebastian Seitner	
C. Jeudy	
Junaid Memon (IPK)	
Katriina Mouhu	
Lukas Kronenberg	lukas.kronenberg@usys.ethz.ch
Alex EUS	

M. Khafif	
Angelo Petrozza	
Michela Janni	michela.janni@ibbr.cnr.it
Alan Gay	abg@aber.ac.uk
Laura Gargiulo	
Rick van de Zedde	
Pierre Lejeune ULIege	
Donald Hobern	
Hund Andreas	andreas.hund@usys.ethz.ch
Ellen	

4. ANNEXES: AGENDA AND PRESENTATIONS

- Link to the presentation material: https://www.phenome-emphasis.fr/content/download/4035/39209/version/1/file/Tools%20for%20variables%20naming%20and%20URI%20generation.pdf

- Agenda :
  - Day1: Friday, April 16, 10:30 - 12:00 CET: Presentation of the interface and models. (Video + in situ demonstration and short discussion), with a reminder of the data model behind it.
  - Day2 : Tuesday, April 20th, 13:00 - 14:30 CET : Hands on practice with input of your own data +
  - Link to the Agenda: https://www.phenome-emphasis.fr/content/download/3985/38882/version/1/file/common%20tool%20for%20variable%20naming%20and%20URI%20generation.pdf