EMPHASIS Training

URI & Variables

Session #1 : Presentation

1. Discussion about URI :

- **Unique vs. un-ambiguous**
  cf. telephone number : you can have 2 mobiles (not unique), but every time someone phones you, it is *you* answering (non-ambiguous).

- **How to make sure an URI is unambiguous if I can design it myself ?**

  => URI is under the authority of a domain name (ex: wur.nl, ipk-gatersleben.de, etc). Under this domain name this is your responsibility to design unambiguous patterns. You cannot have two : inrae.fr/plant12, but there is no problem with a wur.nl/plant12.
  Plant12 is ambiguous, but thanks to the domain name it becomes a non-ambiguous URI.

- **Rename a URI ? :** no, but you can give a new URI and make a re-mapping / re-direction (see https://hal.archives-ouvertes.fr/hal-02390920/document)

2. Discussion about Variables :

- **Do not mix metadata and name**. Only intrinsic quality of the object can be present.

- Every component (entity, charac, unit) has a proper meaning, it exists by itself, however the method can be seen as empty. If we want to give a proper meaning to a variable we should highly detail the method, including some parameters.

- How to model the experimental factors, Genotype, Repetitions, Treatment ?
  - ➢ These variables are slightly different, they are not exactly measures, but experimental design. Thus we model them a bit differently. A factor model is available.

Session #2 : Practice & Exchange

1. Spreadsheet to practice

https://docs.google.com/spreadsheets/d/1x4PANsHecMMdhVo1j2zHeEsAPzG4cjGWdHZe-9exuvE/edit#gid=0

2. Questions

- **What level of detail in method? Include program/supplier? Parameters?**

  => 2 methods should be different enough to allow the definition of two different variables if the method is different enough. Image analysis is not precise enough for a method. When specifying the exact wavelength is probably too much details. (We can use characteristic instead). We can use a Provenance object to specify the parameters of the method and the algorithm used. The variable can also contain some details (but share the same method) if you need to differentiate 2 variables with slightly different methods parameters.

- **Underscores vs Capitalisation** (George Sainsbury - APPF)

  => We adopted a underscore to separate Entity_Characteristic_Method_Unit. If a Method is to be written in multiple words, the Capitalisation (camelcase) is not the best option as URI are case insensitive. hyphen (-) can be used instead to separate words.

- **Environmental variables :** Air as an entity is not precise enough. (Peter Roos)

  => Specified with the scientific object to which the data are assigned. How environmental values are assigned to scientifics objects is a question of modelling and statistical inference.

- **Limit between Entity and Characteristics,** for example here : Plant vs Shoot (Mickael Lamboeuf)

  => We can create entities if the exact definition is lacking, but we should try as much as possible to re-use definitions from ontologies.

- **Link to other ontologies ?** (Andreas Hund)

=> We can create 2 kinds of links :

1- import variables from an ontology, using it's URI. Do not generate a local URI.

2 - make associations between locally defined variables (local URI) and concepts from ontologies. (after the declaration of a variable we can make the link)

- **Trait : Entity + Characteristic**

=> Used in MIAPPE, Crop Ontology, we have improved this model slightly splitting the trait between Entity and Characteristic. That makes more sense for environmental variables.

- **How is the URI format compared to already established ones, for example DOI's ?** (Junaid Memon)

=> The difference is the absence of a central authority (all DOI are generated by the same authority) and the possibility to add semantics in the URI, you can understand rapidly what the ID is. When it is impossible to see what is behind a DOI.

- **How long the URI's can be reserved for certain variables?** (Junaid Memon)

=> Depends on the range of the object. If the object has been generated in a local infrastructure and nobody else is using it, this is fine to have a short life cycle. However if the URI goes online and is possibly used by anyone, you should keep it longer (10 years or more).

- **Very large method names are difficult to use** but we can use the alternative name to use it in graphs or table headers. (Barbro Winkler)

- **Data validation, how to keep the database clean ? :**
    - Priority is for users to keep track of their variables, so one should not be afraid of adding it's variables.
    - In order to differentiate the "standards" from the new additions from the community, we can add details. If the variables is generated by a random user, this is probably not considered a standard, however the community

can consider some variables as references and the administrator can label them as standards.

Links to PHIS Demo:

- URL : http://138.102.159.36:8084/app/
- Login / Pwd : test@opensilex.org / test