

[FIRST SLIDE]

Introduction

Hello and welcome everyone! I am Hannelore Heuer, researcher at the Wageningen Food and Biobased Research Institute in the Netherlands. As the name of our institute suggests, we perform applied research for sustainable innovations in healthy food, fresh food chains and biobased products. We, as the Food Informatics group, digitalize their knowledge and facilitate sharing and linking research output. Currently the world is facing a large population growth, which poses challenges regarding food security. One of the research themes for all of Wageningen University and Research is 'Proteins for life'. Hence, this requires domain experts, governmental organisations, policy makers, you name it, to communicate about proteins.

[SECOND SLIDE]

From this, the need for a Protein Application Ontology arose. We named it 'P-PRO' and use it to enable data sharing between all different parties involved in protein research and application. Now you must be asking yourself 'ontologies that focus on the structure of proteins and their biochemical processes already exist such as PRO and GO, and we have FoodOn, what do we need another protein ontology for?' Well, these ontologies do not really focus on the application of proteins in the industry, and therefore they are not complete to annotate or link all existing data about proteins. They do not always provide the right level of communication about proteins between different disciplines. We have worked on a complementary ontology that can be used by various disciplines and for different purposes such as data annotation and data linking.

PPRO in the web of other ontologies

We made this overview according to the already existing ontologies out there, similar to the FoodOn overview. Our ontology includes expert information about Protein Technology and Human Nutrition and Health, and links to anatomy and taxonomy ontologies, measurement ontologies and to PRO. Researchers from Protein Technology measure protein functionality on specific samples. A big part of the research into Human Nutrition and Health regards digestibility and digestibility measures, and sensory assessments of food. From expert interviews, we drafted the first version of the ontology. We are currently in the process of validating the ontology by interviews with industry domain experts in industry. Additionally, we are working on automatic extraction of protein digestibility information from scientific publications. The extracted terms will then be added to P-PRO to extend it further.

[THIRD SLIDE]

PPRO structure

When we zoom in into the structure of P-PRO, we see that the main concept is the *processed material*. A processed material corresponds to a protein containing food substance that has undergone at least one processing step. The processed material can have a *brand* and a *sample id*. This processed material can only occur in a certain quantity and thus we have connected it to the *om:Quantity* concept.

OM side step

OM refers to the Ontology of Units of Measure, which is one the most comprehensive ontologies to express quantities and units. I have got this example of the diameter of an apple to demonstrate the structure of OM's Quantity class. The diameter of the apple is a quantity, it has a measured value of 10 centimetres, which has a unit in centimetre and a numerical value of 10. The diameter occurs in the phenomenon of

apple. OM is based on several official paper standards, such as: [The Guide for the Use of the International System of Units](#), by the National Institute of Standards and Technology.

PPRO structure

When we go back to our P-PRO ontology, the *raw material*, or *basic food reference* as it is called in FoodOn, contrasts with the processed material, as it is a potential food product that is not yet processed and is not composed of more than one organism. This raw material, just like in FoodOn, comes from a NCBI Taxon *species* which can be further specified with their own *variety* or *cultivar*, and contains (some part of) a plant, fungus or animal.

These materials can be used as an ingredient for a *processingStep*, or they can be the output of a *processingStep*. Processing steps use a specific *method*, one or more *ingredients* and *additives*, a *duration*, *concentration*, *equipment*, *pH*, *temperature*, *material used*, and, last but not least, one or more outputs.

The raw and processed material can have many *observed quantities*, which are defined as OM quantities, such as solubility, viscosity, etc. Since it is important to keep track of how these quantities are measured, a *measurement* concept was added and linked to OM. The observed quantities also include concepts about digestibility, such as the Protein Digestibility Corrected Amino Acid Score (*PDCAAS*) and Digestible Indispensable Amino Acid Score (*DIAAS*). Moreover, the quantities include a *sensory intensity*, which combines a specific *sensory category*, such as 'bland', 'burnt', 'cheesy', with a score for each of these categories.

[FOURTH SLIDE]

Conclusion

I have given you a quick overview of the structure of our Protein Application ontology. This ontology is still a work in progress. PPRO will be available as an open source ontology. We want this ontology to be adopted widely so it can be added to or adapted by the standards already established.

Are there any questions?