Potato Research (2007) 50:357–361
DOI 10.1007/s11540-008-9054-9

The Canon of Potato Science: 36. Potato Ontology

A. J. Haverkort

Published online: 3 June 2008

What is it?

The potato industry – farmers, their suppliers, consultants and procurers – increasingly deals with gathering and processing of large amounts of data. Four main categories of data in arable farming can be distinguished:

- Mandatory registration of data required by legal regulations to obtain a licence to produce. Location and size of the fields and mineral balances may be part of these data as government bodies often require them. Other mandatory data are those regarding food safety – timing of application and dose and type of hazardous chemicals applied – and those required by procurement agents.
- Voluntary registration data are needed if growers volunteer to obtain certificates and labels, as required by the purchaser (e.g., organic) as a license to deliver to wholesale or to the processing industry.
- Data from operations registered by growers for future reference by advisors, for example in decision support systems (DSS). DSS need input based on measurements of soil, air and crop variables, together with a quantitative model or database. Also data from observations in samples of the produce delivered to the processing factory are operational data.
- Generic data, for example data that do not pertain to a specific potato field but still have importance to growers, such as weather conditions, information about market development and financial information on leasing land or labour costs.

The ever increasing amount of data gathered by more growers in more years offers possibilities to add value. Therefore – for interested parties, stakeholders – a common and controlled vocabulary of the potato domain that describes concepts, attributes and the relations between them in a formal way using a standardized knowledge representation language is called for: a potato ontology. The advantage is

A. J. Haverkort (✉)
Plant Research International, Wageningen University and Research Centre,
P.O. Box 16, 6700 AA Wageningen, The Netherlands
e-mail: anton.haverkort@wur.nl

that all possible stakeholders will be able to understand the data expressed by this ontology and that software applications can process them automatically. It will also allow the application of advanced numerical techniques that may help to uncover previously unknown correlations.

Stanford University in the United States of America describes an ontology as a "formal explicit description of concepts in a domain of discourse (classes, sometimes called concepts), properties of each concept describing various features and attributes of the concept (slots, sometimes called roles or properties), and restrictions on slots (facets, sometimes called role restrictions)". An ontology together with a set of individual instances of classes constitutes a knowledge base. An ontology uses a standardized representation language.

To overcome the disadvantages of current practices in data mining, benchmarking and self-learning systems and to benefit from the data collected anywhere, a potato ontology offers the structured approach that is needed. To create a potato ontology one has to:

- First compile and map existing data models, either loosely described or formally defined;
- Analyse these data models for their mutual coherence and consistency;
- Next, propose a unified model that relates all different entities and attributes, in terms of an ontology as defined by the World Wide Web Consortium (W3C).

For potato this may translate into a potato ontology that formally describes:

- "Concepts" or "Classes", e.g., Biocides are a sub-class of Agro-chemicals and has a sub-class of its own Fungicides which in turn is a sub-sub-class of Agrochemicals.
- The "Properties" of the classes, e.g., Agrochemicals are produced synthetically in a factory, biocides are used to protect crops and fungicides to control fungi.
- "Attributes" of the concepts (properties); for example, all agrochemicals have properties such as modes of application and time and dose of application allotting them to slots. "Restrictions" may be that a particular chemical can only be applied with a certain type of equipment, or its application is restricted to a certain period or dose.
- "Instances" in an ontology are individual data such as a particular herbicide treatment with property data: field, time, dose, active ingredient, trademark, mode of application, which equipment operated by whom.
- The standardization language used often is the "Ontology Web Language" (OWL) Protégé developed by Stanford University.

Why is it Important in Potato Science?

The importance of a potato ontology in potato science is that it delivers the environment and definitions needed to advance data management techniques such as storage and retrieval and exploration in potato research and supply chain management. There is an increased continuum between data delivered by practice – as shown above – and those delivered by science such as simulation models, databases (from sciences such as genomics, metabolomics, physiology and agronomy) enabled by automation, decision support systems and search machines. Researchers (but also the industry) potentially (if granted by the data owner) have access to such data and they are then able to develop numerical techniques for *ex-ante* impact assessment of proposed altered scientific or operational practices and to develop and test benchmarking and data mining techniques, query and answer systems and self-learning systems.

Potato - more than most other commodities - is part of a complex industry with specific scientific approaches and solutions. The crop itself is multiplied clonally and so accumulates many pests and diseases. This necessitates a complex seed potato industry. Its produce - the tubers - consists of vulnerable, hard-to-store living material requiring extensive research and development. The potato tuber itself has many qualities that are expressed through the interaction of its genetic make-up (variety), environment and management practices. Consequently there is a wide range of adapted and accepted varieties for the fresh potato market and for processing into starch, frozen and chilled products as well as dry extruded and fried products. This yields (more than for most other commodities) a very extensive and complex database. It is a scientific challenge to shape such a database into an ontology for further value adding in science and practice.

For science it is important to create an ontology as it will allow the development of numerical techniques such as benchmarking, data mining and self-learning systems in one of the most complicated agricultural commodities. The development of concepts, properties and restrictions along rigid hierarchical lines brings much needed structure in seemingly unrelated data that subsequently can be approached in a rational way such as through the use of World Wide Web search engines. The potato ontology will allow scientists to create a query and answer system based on the "competency questions" that various stakeholders have.

Why is it Important for the Potato Industry?

The users of the ontology are all stakeholders of the potato industry. Each group of users has its own competency questions showing why the ontology is important to them:

- The potato grower more easily supplies information to interested parties such as processing companies and auditors of special labels or certificates. The grower has access to all information contained in the database and is the only one to authorize other parties to have access to all or part of it. Improved feed back from the processing industry and consultancy agents that deliver services such as DSS favours the grower. Growers want such specific questions to be answered as those related to irrigation management and risk of nitrogen leaching or less specific questions such as how to manage a seed potato crop in a newly acquired field.
- Authorities want to verify compliance and derive from explorations carried out with the well structured database – based upon the ontology – future policies for agricultural production. Authorities may verify data themselves through audits or may have this carried out by accredited Trusted Third Parties (TTP). Authorities only have access to part of the database they are legally entitled to (e.g.,

compliance to legislation) or are given access to on voluntary basis. An example of a question governments may have is the impact of a policy measure (e.g., restriction on use of water for irrigation) on profitability of certain classes of farms.

- The potato processing industry will want to look into part of the database that is relevant for them. Examples are food safety aspects, such as time and dose of pesticides applied, and how yield and quality were brought about by the individual grower. Benchmarking performance of different growers will allow them to guide them and so improve yield and quality while cutting costs at farm and factory levels.
- Providers of services, especially of decision support systems will want to improve the quality of their advice. When they are able to take all management practices of the grower and of all relevant peers into account they can further fine-tune the advice given.

Scientific Developments

Potato researchers currently employ the Web Ontology Language to express the data items presented above (Protégé 2006) as a modelling tool for the analysis of the terms mentioned in the previous sections. Present scientific developments in potato ontology are the description of concepts in relevant classes. An option being explored presently is to define a super-super class Arable_Farming_Characteristics with super class Potato_Production_Characteristics and class Processing_Potato_Production_Characteristics (along with e.g. Table_Potato_Production_Characteristics, Seed_Potato_Production_Characteristics and Starch_Potato_Production_Characteristics). The ongoing developments concern the subsequent concepts. Sub-classes of Processing_Potato_Production_Characteristics may be Farm_Administration, Field_Characteristics and Crop_Characteristics. The latter then has as sub-sub classes Crop_Organs (with sub-sub-sub class Tuber_Characteristics with slots e.g., Dry_Matter_Concentration and instances e.g., 22.3% (of one of the crops in the last slot). The whole classification is subject of recent and ongoing research. A class (or concept) "Treatments" may be a sub-class of Crop_Management which is a sub-class of Farm_Administration but alternatively may be a sub-class of Field_Characteristics (as they are applied to a field) or a sub-class of Crop_Characteristics (as they are applied to a crop). The scientific advantages and disadvantages of such choices and the effectiveness of a particular built-up related to answering the competency questions of the various stakeholders are part of the present scientific developments. Early interactions with future users in science and the potato industry is part of the current scientific approach to ensure future use.

Further Reading

FAO (2006) AGROVOC Thesaurus, http://www.fao.or/aims/ag_intro.htm

Haverkort AJ, Top JL, Verdenius F (2006) Organizing data in arable farming: towards an ontology of processing potato. Potato Res 49:177–201

Koenderink NJJP, Hulzebos JL, Rijgersberg H, Top JL (2005) Food informatics: sharing food knowledge for research and development. EFITA/WCCA 2005, Vila Real, Portugal, pp 1161–1168

Korotkiy M, Top JL (2006) Onto-SOA: from ontology enables SOA to service enable ontologies. In: Proceedings of an International Conference on Internet and Web Application and Services (ICIW’06), Guadeloupe, pp 392–396

Noy NF, McGuinness DM (2000) Ontology Development 101: A guide to creating your first ontology. http://protege.stanford.edu/publications/ontology_development/ontology101.pdf

Protégé (2006) version December 2005, http://protege.standford.ed/