

The Potato Ontology: Delimitation of the Domain, Modelling Concepts, and Prospects of Performance

A. J. Haverkort · J. L. Top



Received: 3 June 2010 / Accepted: 21 December 2010 /

Published online: 5 February 2011

© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract The ever increasing amount of data gathered by more growers in more years offers possibilities to add value. Therefore—for interested parties and 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 standardised knowledge representation language is being developed: a potato ontology. The advantage is 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. This version of the potato ontology aims at the domain of processing potatoes in a setting of mechanised potato production where growers have access to automated decision support systems and exchange data electronically. This paper describes the procedures to establish such an ontology where competency questions formulated by stakeholders and potential users take a central position. The potato ontology formally describes “Concepts” or “Classes”. The three main classes are those used in crop ecology: Crop, Environment and Management. Classes, e.g., biocides are a subclass of agrochemicals, and in turn have a subclass Fungicides. The ontology also describes the “Properties” of classes, e.g., agrochemicals are produced synthetically in a factory; biocides are used to protect crops and fungicides to control fungi. The ontology also describes the “Attributes” (properties) of the concepts, e.g., all agrochemicals have attributes such as dose and time of application and mode of application. “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. The ontology also features “Instances” which are the individual data such as a particular herbicide treatment with values for field, time, dose, active ingredient, trademark,

A. J. Haverkort (✉)

Plant Research International, Wageningen UR, PO Box 16, 6700 AA Wageningen, The Netherlands
e-mail: anton.haverkort@wur.nl

J. L. Top

VU University Amsterdam, Boelelaan 1081a, 1081 HV Amsterdam, The Netherlands

mode of application, which equipment operated by whom. The standardisation language used is the “Ontology Web Language”.

Keywords Crop performance · Environment · Genotype · Management · Processing potato

Introduction

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 as 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—time, dose, and type of hazardous chemicals applied—and those required by procurement agents
- Voluntary registration data are needed when growers volunteer to obtain certificates and labels, as required by the purchaser (e.g., organic) as a licence 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 from measurements on soil, air or crop, 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, e.g., 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

Modern arable farming is in a rapid transition from applying experience gained through trial and error and oral or written knowledge transfer, to the use of decision support systems. These systems aim at optimising the use of resources and aid in planning production and quality. They do hardly make use of anecdotal and personal experience of the producer but are based on decision support systems with the following elements:

- Recording of crop, soil, or ambient information
- Feeding these data to a quantitative model or database
- Generating advice such as timing, placement, and dose of propagation material or a chemical

These systems and added to them data from mandatory information required by government (as a licence to deliver) and from the processing industry (yield and quality data) lead to a comprehensive database per field that yearly increases in size.

The data per field can be aggregated at various levels such as:

- The farm level to be used in the annual farm management and bookkeeping system
- The series of all years a particular crop was grown in a particular field

- The peer group of all fields with this crop grown in, e.g., the same or a series of years, in similar or different soil types in the town area or in many areas

All types of data may be processed by software systems built for different purposes, which initially did not necessarily intend to communicate.

The emergence of knowledge representation languages allowed the retrieval and transfer of information placed in the public domain or within pre-arranged parties such as producer and processor. For example, Edi-teelt (Haverkort et al., 2006) is a Dutch system by which growers send their crop registration data in a standardised internet way to, e.g., the Inspection Service, the Food Safety Certifying agent, and so forth. Meanwhile, the focus shifts towards the definition of concepts, independent from particular software systems to allow better searching to interpretation of knowledge. This resulted in the development of *ontologies*. An ontology is a controlled and shared vocabulary that describes concepts and the relations between them in a formal way, and has a grammar for using the vocabulary terms to express something meaningful within a specified domain of interest. Having adequately defined ontologies benefits both human interpretation and supporting machine processing.

An example of a related, existing ontology is the “Plant Ontology” (Plant Ontology Consortium, 2010) that consists of two parts. One treats the plant structure: “A controlled vocabulary of botanical terms describing morphological and anatomical structures representing organ, tissue and cell types, and their relationships. Examples are gametophyte, parenchyma, guard cell, etc.” The other part treats the growth and developmental stages: “A controlled vocabulary of terms describing (1) whole plant growth stages and (2) plant structure developmental stages. Examples are seedling growth, rosette growth, leaf development stages, embryo development stages, flower development stages, etc.” The Plant Ontology describes botanical concepts and their hierarchical structure and mainly aims at facilitating cross database querying.

Another example (Fig. 1) of an ontology relevant for agriculture is AGROVOC (FAO, 2010) created and maintained by the Food and Agriculture Organization of the United Nations. It is a multilingual vocabulary and covers the terminology of agriculture and related domains. The tool is in the public domain and has searching and browsing functionalities.

Here, we present a potato ontology—a case of an arable farming ontology—as an example of a single information model that eliminates confusions and clarifies unclear notions. A widely accepted standard for modelling the domain of arable farming helps to improve communication between consultants, researchers, farmers

Fig. 1 Part of the AGROVOC Thesaurus (FAO, 2010)



and government agents because it provides a shared vocabulary. Even more interestingly, an ontology in this area facilitates unambiguous communication between computers and supports automation of processes such as data mining, access policy to data by, e.g., governmental services, and processors (thus reducing the administrative workload) and providers of decision support systems that are individually owned by farmers. One of the greatest challenges in knowledge management is to deliver systems that can learn automatically from experience. Computer systems exist that can extract new knowledge from empirical data. This knowledge can actually lead to solutions for problems that could not have been solved before (Verdenius, 2005). It is the challenge to combine existing expertise and scientific knowledge with the new knowledge that is obtained from data (Verdenius and Top, 1998).

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 standardised knowledge representation language is called for; a potato ontology. The advantage is 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 enable the application of advanced numerical techniques that may help to uncover previously unknown correlations.

Our ambition is 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 make a potato ontology, we had to:

- First compile and map existing data models, either loosely described or formally defined
- Analyse them 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)

Stanford University 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 constitutes a knowledge base. An ontology uses a standardised representation language, we used Ontology WEB Language Resource Description Framework Schema as defined by the W3C consortium.

In a preceding paper (Haverkort et al. 2006), we provided an overview of existing data sources and data items. We described the rationale of an ontology for arable farming. In particular, we focused on data pertaining to potato production for processing potato in the Netherlands as an example of a field-grown crop destined for processing. There we compiled and mapped existing data models (either loosely described or formally defined) and we analysed them for their mutual coherence and consistency. In the present paper, we propose a unified model that relates all different entities and attributes, in terms of an ontology as defined by W3C and of which the plant ontology and AGROVOC are examples. The unifying model of the potato

ontology presented here is based on the agro-ecological principle that (1) crop performance is determined by (2) genotype, (3) environment, and (4) management. These form the four elements along which the ontology is structured. We show the domain that is covered by the ontology based on the types of questions it should help answer. Examples of its application do not exist yet as no study has been carried out yet that filled (part of) the ontology with relevant instances. With the increasing amount of data becoming available and prospects of capitalising on them increasing, it seems a matter of time before the need will be felt and then the presented ontology may serve as the structure needed for the desired functionalities such as an effective query mechanism in data mining. We have taken the situation in the Netherlands as an example, but most concepts and relations are similar in other countries with fully mechanised production systems. The unifying principle in crop production ecology approach that crop production is a result of interaction of genotype, environment, and management is used here to help give structure to the ontology. We followed the procedures to set up the ontology as shown by Noy and McGuinness (2001) and in the following will describe the domain, supply competency questions, enumerate important terms and structure classes and subclasses with the attributes.

The Domain and Scope of the Potato Ontology

The Domain Covered by the Potato Ontology

The domain we have chosen is the production of potato for processing into French fries, crisps, flakes, or starch. It covers the whole range of activities from field selection and preparation, selection and planting of seed tubers, crop management, harvest and storage until the potatoes are collected by the processors, and yield and quality are determined by them or independent assessors. Spatial and temporal aspects of these managerial activities are part of the domain as well as the recording of generic and field and crop specific data related to them. These pertain to a specific potato field and only where pertinent whole farm aspects such as crop rotation are included. The following available databases are part of the ontology:

- Generic data such as weather and most of the soil data
- Data collected from lists such as the National List of Varieties and Kwantitatieve Informatie (Dekker 2006)
- Mandatory recorded data by the grower needed by authorities and processors (licences to deliver)
- Data recorded by the grower such as field history, seed variety and quality and when activities were carried out with what kind of equipment and with which dose
- Recorded data collected by the grower or by scouts of service providers as input for DSS such as counts of trapped pests, soil mineral content, and stages of crop growth
- Output of DSS such as recommended time, type, mode, and dose of an application of chemicals such as fertilisers or biocides
- Data feedback from the processor regarding yield, tare, processing quality to reduction due to defects

For What Types of Questions the Information in the Ontology Should Provide Answers?

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 authorise other parties' access to all or part of it. Improved feedback from the processing industry to consultancy agents that deliver services such as DSS favours the grower. Growers want specific questions answered such as questions 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 to 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 the profit of certain classes of growers
- The potato processing industry will want to look into part of the database that is relevant for them: food safety aspects such as time and dose of pesticides applied but probably even more in how yield and quality was brought about by the 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

Competency Questions

The potato ontology we modelled potentially should be used by the above group of users, each questioning the knowledge base with specific questions for specific purposes. To determine the scope of the potato ontology, we first sketched a list of potential interested parties and for each of them a few questions that the potato knowledge base based on the ontology should be able to answer, competency questions (Gruninger and Fox 1995). These questions will serve as the litmus test later: Does the ontology contain enough information to answer these types of questions? Do the answers require a particular level of detail or representation of a particular area? These competency questions are just a sketch and are not exhaustive. They are only illustrated for four groups of potential users: growers, service providers, processing industry, and certifying agencies.

Potato growers are interested in reducing costs and increasing yield and quality (i.e., crop value). Learning from previous experience is crucial, especially in present times, with increased scale of production. A grower less frequently knows the field planted with potatoes as the land may be rented or the scale of production is such that too many fields are involved. Examples of competency questions potato growers may have are the following:

- Which combination of inputs gives the highest yield and dry matter concentration (i.e., market value?)
- What is the best blight control for early potato varieties?
- Which management options do I have for processing potatoes once planted?
- Which combination of nitrogen fertilisation and irrigation is optimal (variety and soil type specific) to arrive at the lowest concentration of reducing sugars?
- How to achieve multiple goals such as saving on fertiliser while still achieving commercially acceptable yields?

It is obvious that the knowledge base is more relevant when it contains more fields (including those of as many comparable peers as possible) and years.

Extension services and firms providing consultancies and delivering decision support systems (e.g., for timing and dose of biocides, fertilisers, and water) are equally interested in field-specific past performance of crops and responses to crop and variety, weather, and management. This allows them to better fine-tune the decisions they make—especially when one is allowed access to decision support modules supplied by a different service provider. A laboratory advising on supplemental nitrogen dressing would benefit to know the advise on irrigation as to fine-tune the advise regarding risk of over-irrigation (leading to leaching of nitrogen) or under-irrigating (reducing growth hence the risk of over-fertilisation).

Competency questions from farmers' service providers:

- Do farmers with greater amount of data collected (e.g., because they have used more DSS) have higher yields and/or better quality?
- How can we assist growers in data mining and development of self-learning systems?
- How can we make different DSSs communicate as to improve them through synergism?

Processing companies in the Netherlands alone process about 3.5 million tonnes of ware potato into frozen and fried products. Another 2.5 million tonnes is processed into starch (and derivatives) and protein. The same holds for other potato-producing and -processing countries in Europe and North America. They usually have a number of major concerns:

- Low-cost raw material (potato tubers procured) at competing prices meeting specifications as close as possible
- High-quality raw material (expressed as recovery, i.e., quantity of finished product per tonne of raw material processed). In practice, the factors mainly influencing recovery usually are high dry matter and low reducing sugars concentrations, tuber size, and shape and amount of defects

- Which key factors determine sugar concentration? And how do they interact?
- Inconsistent raw material requires a constant fine-tuning (leading to costs and reduced recovery) of the factory processing conditions to meet the specifications of the finished product

Examples of competency questions of the processing industry are:

- Which factor most strongly reduces seasonal (regional) and spatial variation?
- What is the relation between field size and yield and quality (consistency)?
- Which potato variety keeps the highest dry matter concentration on lighter soils?
- Which variety–area where grown–management combination leads to the lowest incidence of defects?
- How to disentangle managerial and environmental factors that determine tuber shape in a population of 600 growers?
- How to reduce cost of raw material production as to obtain reasonable contracts with growers?
- Raw material should be produced in a sustainable way regarding people (healthy food produced under save conditions), planet (optimal use of resources and minimal emissions of chemicals), and profit (best return on investment). Companies are looking for indicators of sustainability (Haverkort et al., 2009) and wonder if an ontology helps

Certifying agencies are accredited to give a farm or a product a label certifying that the producer complies with the requirement of the label (Haverkort 2008). These may be labels issued by a processing or retail company, by environmental non-governmental organisations or related to organic farming. They may be related to food safety, environmental friendliness of production, or both. Regarding tracing and tracking of the labelled produce, their interest lies in access to the farm's database. Examples of competency questions they may have are:

- When was a certain chemical product applied, what was the dose and with what kind of equipment and what likelihood of residue is there?
- What chemical–time–dose–equipment–weather combinations lead to what kind of risks of residues? This in view of reducing or decreasing sampling time or adjustment of requirements
- Can the ontology contribute to the establishment of sustainability indicators and their values?

Researchers in close cooperation with leaders of the potato industry may have questions at strategic levels that may result in a research agenda. Questions arising here are, e.g.,:

- Which are the main determinant factors (in decreasing order of importance) describing incompetent management?
- How do these interact?
- Which variety is most susceptible to incompetent management?
- How to disentangle managerial and environmental factors that determine tuber shape in the population of all contracted growers?
- Which management options exist to adapt to various scenarios of climate change?

Systems Analysis of Processing Potato Production

Before the modelling of concepts took place, we first systematically analysed the potato crop. Therefore, we first enumerated the important terms in the ontology by describing the supply and production chain and subsequently described the domain in a logical order used by crop modellers (genotype \times environment \times management interactions).

Enumeration of Important Terms in the Ontology

Before making the ontology, we first made a list of all terms we would like either to make statements about or to explain to future users. What follows here is by far not exhaustive but shows the scope of the domain. It draws the attention to concepts and properties; about every other word *in italics* in what follows below (example first part only) is a term that will show up in the potato ontology.

Potato is used as *seed potato*, as *fresh table potato* after boiling, or *processed* such as fried in oil (French fries) and frozen and consumed at quick service restaurants or into *crisps*, *flakes* or *starch*. The *potato crop* is grown on a *field* after *soil preparation* in a specific area with its *characteristic soil* (percentage *clay*, *organic matter*, *mineral content*, and *rooting depth*) and *weather* (*radiation*, *temperature*, *precipitation*, *evaporation*) properties. *Seed tubers* of a certain *variety* with genetic *varietal properties* (*earliness*, *skin and flesh colour*, and *resistance to pests and diseases*), *seed size*, *physiologic* and *health* specifications are planted at a predetermined *planting pattern* and *depth*. After *emergence*, the crop may be separated in *above-ground biomass* (*stems and leaves*) and *below-ground biomass* (*roots and tubers*). The tubers have a *quality: mass*, a *size* and properties such as a *length–width ratio*, *dry matter* and *sugar* concentrations, and *defects* such as *black spot* and *hollow hearts*.

The grower needs equipment such as tractors, a plough, a harrow, a potato planter, a ridging machine a harvesting machine, and conveyor belts for store loading. The grower also needs stores for machinery and potatoes. During crop growth, the grower needs to manage the crop to assure proper preplanting and supplemental fertilisation with minerals (especially nitrogen, potassium, and phosphorus) and supply the crop with adequate water through irrigation. The crop and or seed potatoes or potatoes in storage may be affected by yield reducing factors of biotic (pests, diseases, and weeds) or abiotic (tare, contamination) factors, the nature, species, incidence, sampling date, and population density need to be established as well as control measures.

To optimise the use of inputs the grower often is assisted by DSS that measure relevant ambient and crop factors, and derive a time, dose, and type of chemical application advice. DSSs are available for control of pests (aphids, Colorado beetle, nematodes), weeds, and diseases. For all management activities, including haulm killing, lifting, store loading, and grading the grower has equipment available. Equipment is of a certain type and age, has a date when last tested, and a known mode of action.

Upon delivery of the potatoes to the factory, various data flows come into existence such as those pertaining to payment (quantity of all potatoes delivered

within agreed-upon grades minus tare and quality such as dry matter concentration, size and defects), licence to deliver (label or certificate), and information to government agencies such as related to mineral balances, use of biocides, and cost–benefit analysis for bookkeeping and tax purposes.

Performance, Genotype, Environment, and Management

Crop production ecologists (e.g., Hammer et al. 2002) define that potato yield and quality (performance, P) are a result or are determined by the crop type and variety (genotype G) that is grown, the environment (E) in which the crop is grown, and by the crop management (M)

$$P = G \times E \times M$$

Performance

The crop performance at periodic or final harvests is the observed crop dry matter production, the distribution of the dry matter over the various organs and their properties such as stem length, rooting depth, leaf area and tuber yield. The latter may be graded in various quality grades such as fodder (e.g., smaller than 35 mm), fresh market (35–55 mm), and French fries grades (over 55 mm). Tuber shape (length–width ratio) and dry matter and reducing sugar concentrations as well as the proportion of the tubers suffering from (a)biotic defects are quantitative indicators of performance. In seed potato the proportion tubers infected with certain pests or diseases, beside size, is the most important indicator of performance.

Sources of performance data may be the grower database, feedback from the processor or trading company the tubers were sold to or—when it concerns the production of seed potato—the proportion of, e.g., tubers infected with *Potato virus Y*.

Figure 2 shows part of the ontology pertaining to the performance part of the potato domain related to the quality of the harvested produce.

Genotype

Of the three groups of factors leading to performance (genotype, environment, and management), genotype is often considered as being the most important one as it determines where and when (environment) the crop is grown, and how it should be treated (management).

The crop (genetic make-up) among others is described by its:

- Type (e.g., starch, crisps, fresh table)
- Relative earliness or lateness
- Variety dependent quality determinants (colour, shape, mealiness...)
- Relative levels of tolerance and resistance to biotic and abiotic influences such as to late blight, drought...

Figure 3 illustrates part of the ontology related to resistance of genotypes to diseases. The main source of information regarding the genotype is the recommended list of

Fig. 2 Illustration of part of the ontology describing the tuber quality of crop performance (non-exhaustive)



varieties that is produced in many countries and gives description of the variety characteristics that are mainly (but not solely) determined by its genotype.

Environment

The environment (E) or field where a particular potato crop is grown is described by its:

- Location as may be shown by its geographical vectors, field size, a map, a photograph...
- Soil characteristics such as granular composition, pH...
- Climate as represented by long-term weather data such a maximum and minimum temperature, rainfall...
- History as described by previous crops such as wheat, sugar beets, and the frequency with which each crop is grown on the particular field

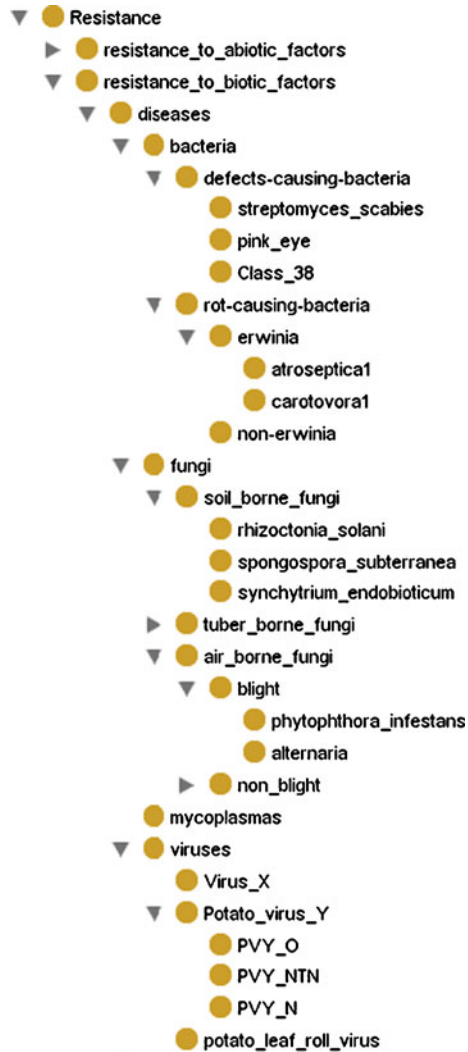
Figure 4 shows part of the ontology of the environment part of the potato domain treating the climate.

Management

The management (M) of the potato crop is described by its:

- Processes such as work plans, bookkeeping, certification, decision support systems...

Fig. 3 Part of the ontology related to resistance of genotypes to diseases (non-exhaustive)



- Equipment such as machinery for tillage, planting, irrigation, nursing, harvesting...
- Buildings such as stores of potato and machinery...
- Material such as biocides, fertiliser, and water

Figure 5 shows part of the ontology of the potato domain that describes management, more specifically the part dealing with labelling of the produce.

Definition of the Classes, Class Hierarchy, Slots, and Instances

Our potato ontology is further explained in Fig. 6 and is illustrated with a screen shot from Protégé in Fig. 7. It describes concepts or classes, their properties, their attributes, and eventually contains instances. After having enumerated most relevant

Fig. 4 Illustration of part of the ontology describing the environment (climate) (non-exhaustive)

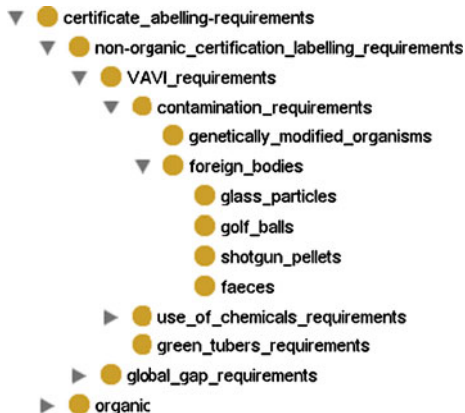


terms and having given them a place in the $P=G \times E \times M$ framework, we defined classes (concepts) and the hierarchy between classes. We followed a topdown development process and started with the definition of the most general concepts in the potato domain and subsequently specialised the concepts as shown in Figs 2, 3, 4 and 5. Subsequently, we defined properties and their delimitations (restrictions).

Delimitation of the Present Scope of the Potato Ontology

We restricted the scope of the potato ontology by, e.g., not including the parents of the varieties and their characteristics although such pedigree (van Berloo et al. 2007) is readily available. Nor did we include the models upon which the decision support systems are based although most of them are found in the public domain. The procedures of determining the quality at processors’ factories are not included either.

Fig. 5 Illustration of part of the ontology describing management (labelling)



If a specific use of the potato ontology would require such addition, the structure allows it.

Potential Deployment of the Potato Ontology

The future users of the potato ontology are stakeholders of the various links of the potato supply chain and their contributors. Figure 8 schematically represents them and shows how each may have access to part of the database described by the ontology. Here, we assume that ultimately the database is owned by the potato producer him or herself but use and maintenance of the ontology is with third parties (text below mainly refers to and partly re-written from Haverkort, 2008).

A potato ontology is a formal description of

- 1: Concepts or classes in the potato ($P = G \times E \times M$) domain, e.g.,
 - `Potato_cyst_nematodes` (sub-sub-sub class) are a kind of `nematodes` (sub-sub class) is a kind of `pest` (sub class) is a kind of `yield_reducing_factor` (class)
 - `French_fries_potatoes` are kind of `processing_potato` are a kind of `potato`
- 2: Classes have slots with intrinsic, extrinsic properties, parts or relationships, e.g.,
 - the class `seed_potato` has slots defined by intrinsic (measurable) properties such as `size`, `certification_level`, `age`
 - the class `seed_potato` may also have extrinsic properties such as the name of the variety, grown on a specific soil type or by `seed_grower_number_12345`.
 - A slot may be a part of a structured object, e.g., `harrowing` is a slot of the class `soil_preparation` where several actions follow one another
 - A slot may be defined by a relationship such as varieties from the same breeder
- 3: Attributes may be restricted in terms of cardinality or value type. The property-type defining attributes may have a single cardinality (colour, variety) or have multiple cardinality (`range_of_seed_sizes`, `types_of_potatoes_grown_by_farmer_12345`). Attributes may have the following value types
 - String like a name or a description
 - A number such as in `seed_price` or `fungicide_concentration`
 - Boolean slots can be answered with yes or no such as in the property `pre-sprouted`
 - Enumerated slots specify a list of specific allowed values for the slot such as `earliness` slot can take the values: `very_early`, `early`, `mid_early`, `mid_late`, `late`, `very_late` or `seed_size`: `between_28mm_and_35mm` and `between_35mm_and_45mm`
4. Creating instances consists of choosing a class, creating individual instances and filling in the slot values. Examples:
 - Class `seed_potato`, instance `seed_lot_planted_on_parcel_road_side` with slot values: Variety `Désirée`, grade 35-45 mm, health status `certificate A`, produced by farmer 12345, price € 550/t
 - Class `machinery`, instance `potato_planter` with slot values: trademark `Högland`, type `4RS`, purchased_in_2006, price €4600, `four_row-system`

Fig. 6 Schematic representation of potato ontology construction

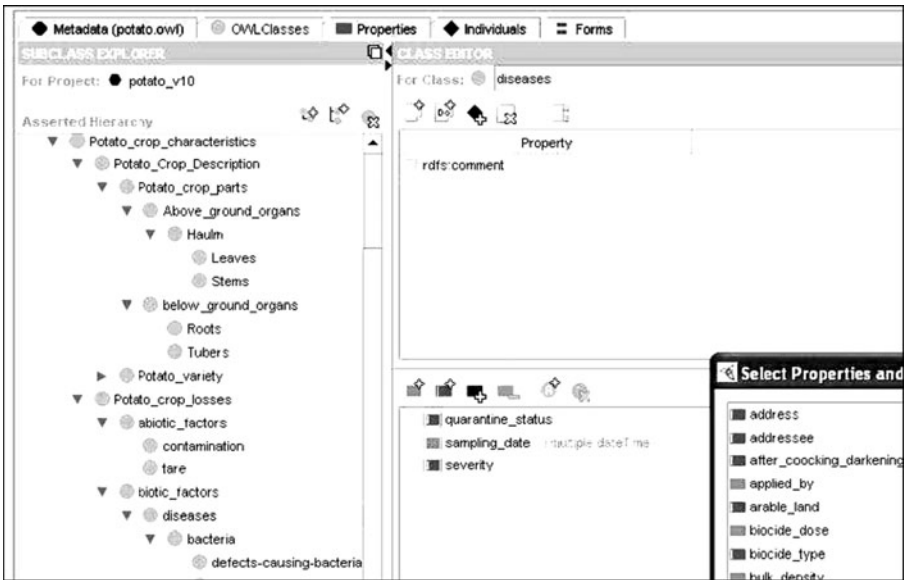
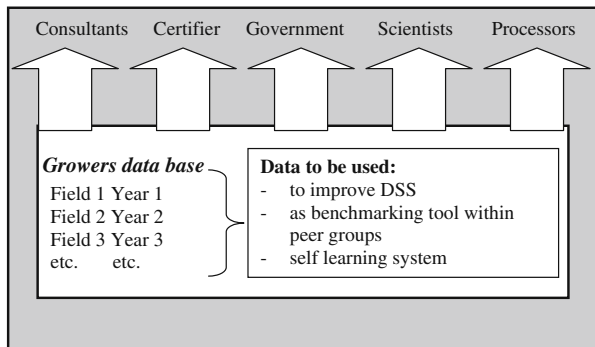


Fig. 7 Screenshot of the modelled potato ontology in Protégé (2006)

- The potato grower interest in the ontology is to more easily and coherently exchange data to interested parties such as processing companies and auditors of special labels or certificates (e.g., GlobalGAP 2010). The grower has access or should have to all information contained in the database and the grower is the only one to grant other parties access to all or parts of it. The growers benefit from improved feedback from the processing industry to consultancy agents that deliver services such as DSS favouring the grower’s management
- Potato producer’s organisations aim to intelligently improve their services. They can direct research and development funds to areas shown—as answers from competency questions—to have the quickest to highest impact as shown through ex-ante assessment using the database
- Authorities that want to verify compliance with regulations. They may derive future policies for agricultural production from the well-structured database provided by the ontology. Authorities may verify data themselves through audits

Fig. 8 Schematic representation of the future users of the potato ontology



- or may have this carried out by accredited TTP. Authorities only have access to part of the database they are legally entitled to (compliance to legislation) or are given access to on voluntary basis by the grower
- The potato-processing industry aims at producing safe food of good quality, grown in an environmentally friendly way at competitive prices. They want to look into part of the database that is relevant for them: food safety aspects such as time and dose of pesticides applied but probably even more in how yield and quality was brought about by the grower as to allow the highest recovery (grammes of finished product per kilogramme raw material procured) in the factory. Benchmarking the performance of different growers between each other and with simulation model calculations will guide them and so enhance yield and quality while cutting costs at farm and factory levels
 - Providers of goods such as seed potatoes, chemical fertilisers, and biocides when analysing past performance may offer better products such as the quality of the progeny (variety characteristics such as resistance to prevailing pests and diseases, expected quality performance to best meet processing specifications) and chemicals
 - Service providers, especially those of decision support systems assisting growers in making decisions on timing and dosage of applications of chemicals will improve the quality of the advice when taking all management practices of the grower into account. When benchmarking it across all growers' fields in their database—with growers consent—and combining it with simulation model calculations will still further fine-tune the advice given

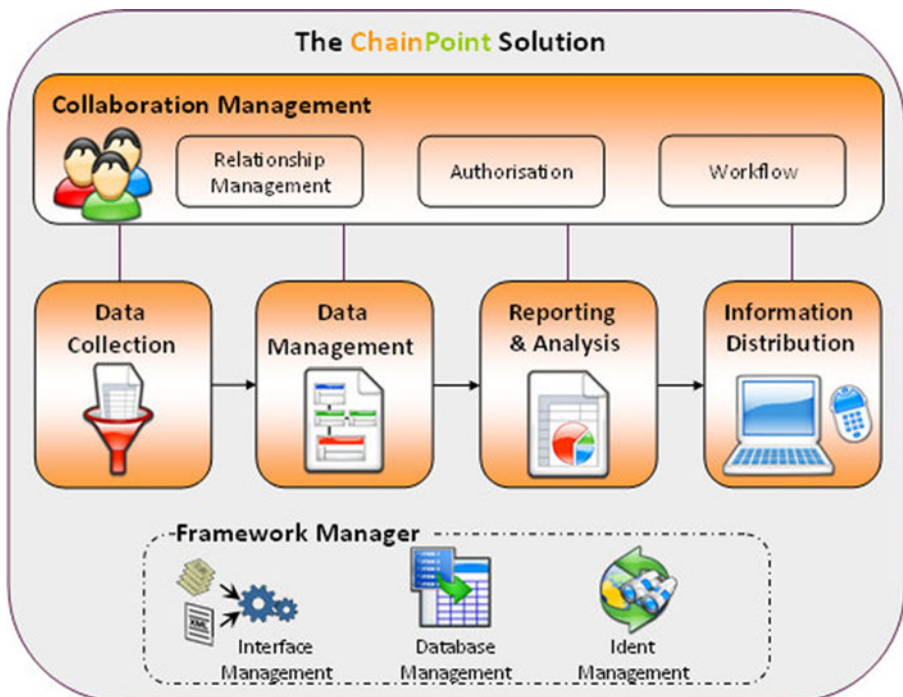


Fig. 9 Example of a service provider showing benefits of integrated data management (source: Chainfood, 2010)

- Scientists have a special interest in the ontology to further develop it and to continuously interact with growers and growers with them. Scientist will want to use the database constructed following the ontology as tool to ex-ante assess impact of the proposed altered management practices (e.g., how much will the water use efficiency increase when water availability is reduced) and also to use it to develop and tests numerical techniques such as benchmarking, query-and-answer analysis, and self-learning systems

The ontology needs to be managed and maintained to assure that all parties continuously will benefit from the aspects shown above. There are several parties amenable to take the lead in this effort. Edi-teelt (Haverkort et al., 2006) is one of them. Alternatives could be one of the leading agricultural information, communication or technology companies, or a consortium of a combination yet to be identified. Potentially also large potato processing companies may be interested in extending their potato database using the structure that the Potato Ontology is offering. When exploring the internet, many information and communication companies offer services to businesses dealing with very large amounts of data. Such companies are able to assist businesses with generically managed data flows (Fig. 9) and are currently investigating possibilities to invest in ontologies.

Figure 9 schematically shows the act of data collection, its management and analysis, and the distribution of information derived from the data. Especially the data management and analysis parts of this set-up may benefit from the definitions to structure that an ontology offers.

Final Remarks

Distributed knowledge bases employing the potato ontology can act as a single knowledge source available to all. Proper authorization mechanisms will ensure that proprietary information, mostly at the level of instances in the knowledge base, are protected. The structure of the shared knowledge base follows the general structure $\text{performance} = \text{genotype} \times \text{environment} \times \text{management}$, giving the ontology a transparent structure. Theoretically, in the not-too-far future, when a potato grower buys an additional parcel of land, he or she will be as much interested in its history as in the physical area purchased. The potato ontology will allow the grower to derive from it the optimal management practices for potato crops grown in different rotations, for different purposes under varying options of management. By articulating the type of crop the grower wants to grow on this plot (ware potato, seed potato, starch potato, early or main crop), the queries posted to the potato ontology supported knowledge base will supply answers as to optimise genotype and management interventions.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Chainfood (2010) Integrated data management system. Available at: <http://www.chainfood.com/chainpoint>. Accessed 16 Dec 2010
- Dekker WA (2006) Kwantitatieve informatie, akkerbouw en vollegrondsgroententeelt (In Dutch, quantitative information for arable farming and field production of vegetables). PPO, Lelystad
- FAO (2010) AGROVOC Thesaurus. Available at: <http://aims.fao.org/website/AGROVOC-Thesaurus/sub>. AGROVOC Thesaurus. Accessed 16 Dec 2010
- GlobalGAP (2010) Website: <http://www.globalgap.org/>. Accessed 20 Jan 2011
- Gruninger M, Fox MS (1995) Methodology for the design and evaluation of ontologies. In: Proceedings of the workshop on basic ontological issues in knowledge sharing, Montreal, April 1995, pp 1–10
- Hammer G, Kropff MJ, Sinclair TR, Porter JR (2002) Future contributions of crop modelling—from heuristics to supporting decision-making to understanding genetic regulation and aiding crop improvement. *Eur J Agron* 18:15–31
- Haverkort AJ, Top JL, Verdenius F (2006) Organizing data in arable farming: towards an ontology of processing potato. *Potato Res* 49:177–201
- Haverkort AJ (2008) The canon of potato science: potato ontology. *Potato Res* 50:399–402
- Haverkort AJ, Bindraban PS, Conijn S, de Ruijter FJ (2009) A crop production ecology approach of sustainable biomass production for food, feed and fuel. *Outlook Agric* 38:249–258
- Noy NF, McGuinness DL (2001) A guide to creating your first ontology. Available at: <http://www.ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noymcguinness-abstract.html>
- Plant Ontology Consortium (2010) The plant ontology. Available at: <http://www.plantontology.org/>. Accessed 16 Dec 2010
- Protégé (2006) Protégé. Available at: <http://protege.stanford.edu/>, version December 2005. Accessed on 16 Dec 2010
- Van Berloo R, Hutten RCB, Van Eck HJ, Visser RGF (2007) An online potato pedigree database resource. *Potato Res* 50:45–57
- Verdenius F (2005) Methodological aspects of designing induction-based applications, PhD thesis, University of Amsterdam. Available at: <http://dare.uva.nl/record/146542>. Accessed on 16 Dec 2010
- Verdenius F, Top JL (1998) Case based modelling of dynamic systems. *Acta Horticulturae* 476:279–288