Potato Research (2006) 49:177–201
DOI 10.1007/s11540-006-9016-z

Organizing Data in Arable Farming: Towards an Ontology of Processing Potato

A. J. Haverkort · J. L. Top · F. Verdenius

Received: 31 January 2006 / Accepted: 11 December 2006 /
Published online: 5 April 2007

Abstract Arable farmers and their suppliers, consultants and procurers are increasingly dealing with gathering and processing of large amounts of data. Data sources are related to mandatory and voluntary registration (certification, tracing and tracking, quality control). Besides data collected for registration purposes, decision support systems for strategic, tactical and operational tasks yield enormous amounts of mainly digital information. Data of similar nature but with often varying definitions are collected and processed separately for different purposes. This paper describes for an important arable crop – the processing potato – which data requirements and flows exist at present and how they could possibly be described in a unifying ontology. An ontology describes the concepts, attributes and relations in a specific knowledge domain using a standardized representation language. Important concepts in this domain are for example crop, parcel, soil, treatment and farm. The ontology – once elaborated – will reduce the overlap between information models and helps to overcome the problem of data definition and representation. It is a key element for the development of systems that can automatically learn either with the help of expert knowledge or through adequate numerical techniques.

Keywords certificate · crop growth model · database management · decision support system · ontology · potato · registration · self-learning system

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

J. L. Top
Vrije Universiteit Amsterdam, De Boelelaan 10081a, 1081 HV Amsterdam, The Netherlands
e-mail: jan.top@wur.nl

J. L. Top · F. Verdenius
Agrotechnology and Food Sciences Group, Wageningen University and Research Centre, P.O. Box 17,
6700 AA Wageningen, The Netherlands
e-mail: floor.verdenius@wur.nl

Introduction

Primary producers of agricultural produce generate increasing quantities of data. These data refer to various entities and serve multiple purposes. Some data are needed to comply with governmental regulations; others are used to optimise production and processing conditions. Data may refer to meteorological conditions, soil characteristics, irrigation schemes or for example supplemental nitrogen dressings. Other data describe the crop that has been grown on a specific field for a specific year, including information on biocide application timing and dose, seed rate and time of planting and harvesting. These data originate from various sources such as material suppliers, specialised data providers (e.g., for meteorological data, soil data), and growers (e.g., seed rate, time of planting and harvesting) and processors.

We distinguish four main categories of data in arable farming:

– *Mandatory registration* required by legal regulations. Without collecting and providing this mandatory data, a grower would have no licence to produce. This information includes mineral balances, quarantine actions and crop protection services. Governmental subsidies such as those related to fallow and commodities subjected to quota (starch and sugar) also require registrations. Moreover, indicating the exact location of a field is mandatory in The Netherlands where the co-ordinates of each field are stored in the database Base Registration of Parcels (BRP) at the Ministry of Agriculture.
– *Voluntary registration* is needed if growers volunteer to obtain certificates and labels, e.g., EUREPGAP (Euro Retailer Produce Good Agricultural Practices). Several protocols for potatoes, vegetables and fruits exist, see for example General Regulations Fruit and Vegetables Cologne (2004). Relevant for potato: General regulations Fruit and vegetables Cologne, Germany FoodPlus GmbH, version 2.1. October 2004. In addition to being used for food safety measures, this certificate also pays attention to environmental and labour related conditions.
– *Operational registration* by growers for future reference and by advisors, and for example in decision support systems (DSS), examples of which are given in "Materials and Methods". These systems need input from measurements on soil, air or crop, together with a quantitative model or database. The output of a DSS can for example be timing and dose of resource applications.
– *Generic data* e.g., data that do not pertain to a specific field but still have importance to growers, such as weather conditions, information about market development and financial information on leasing land or labour costs (KWIN, Dekkers 2002).

These data serve multiple purposes. Growers, consultants and extension services are interested in data to learn how to optimise the on-farm performance. They use the data to develop and enhance their expertise, and to construct and validate the models underlying decision support systems. The processing industry is interested in product information, presently often obtained by systems for tracking and tracing. Traders and processing industries provide feedback information to the grower, such as information on yield and quality of the materials they supplied. Governmental bodies finally request data to assess compliance with rules and regulations. Thus, as a consequence of tighter requirements for farmers, data collection has become an important aspect of farming.

Beside direct use of data in software applications, they also serve the indirect purpose of obtaining insight. An obvious benefit of gathering data from a large group of growers for the potato processing industry would be to learn from these data and to improve future production performance. Typical applications may be benchmarking, allowing producers to measure their performance, or data mining, possibly leading to new understanding of the mechanisms underlying arable farming. However, the results of both benchmarking and data mining are not always easy to interpret.

In the case of benchmarking, growers evaluate their yields relative to a peer group of growers with similar characteristics. These characteristics may include various farm properties and management actions. If the comparison results in totally different outcomes, depending on the characteristics studied, it will hardly be possible to attribute the obtained yield, quality and financial gains to (combinations of) locations. Interpretation of such benchmarking results is not trivial.

An obvious advantage of previously gathered data by a large group of growers for the potato processing industry would be learning from them to improve future performance (Wolfert et al. 2005). Benchmarking and datamining are being tried but their results are not easily interpreted. When growers find their yields in a certain part of a distribution within their peer group and the various farm properties, management actions and use of inputs in numerous other frequency distributions but at different places within such distributions it is hardly possible to attribute obtained yield, quality and financial gain to (combinations of) location in the frequency distributions. Many relationships are to be distrusted e.g. growers that apply more phosphorus tend to have lower yields. The conclusion possible should not read "apply less phosphorus" because probably the heavy users are situated on more phosphorus fixating soils. Registration of the adequate soil characteristics is than needed for a proper benchmarking advice.

Knowledge creation from a field database may also be achieved through a self-learning system. Figure 1 shows the requirements and working of such a system. Data are collected annually on a field and stored in a database. With the aid of quantitative approaches (statistical and mechanistic models) the data are interpreted using yield-gap analysis and other techniques. Suggestions are generated for time, site and dose of inputs. Before being

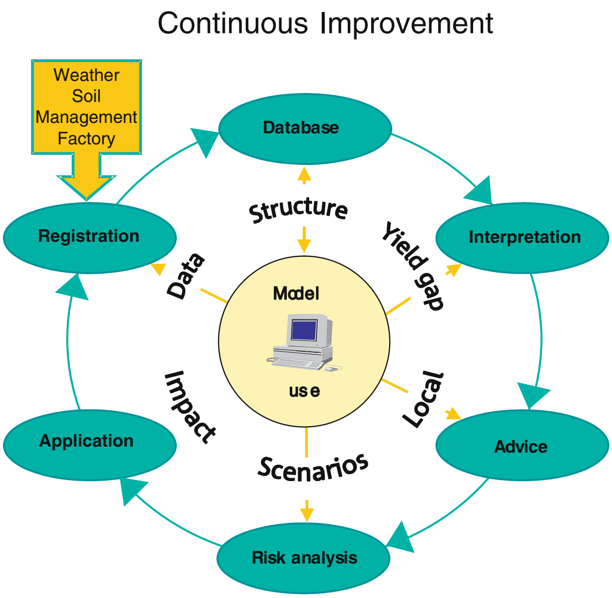

Fig. 1 Schematic representation of a self learning system making use of automated generation of knowledge (from MacKerron and Haverkort 2004)

implemented, this advice is tested in scenarios with long-term average weather data for robustness and risk of failure (too many or too few sprays, over or under irrigation and over or under nitrogen fertilisation). Once the recommendation has been implemented the response is recorded for subsequent years. The database and subsequent recommendations are adjusted accordingly. In this way a continuous learning cycle is created between farmers, extension agencies collecting and interpreting the data, and science. All three parties exchange data and views and improve their performance while contributing to the system.

Data collection for benchmarking or data mining involves numerous data sources, serving different goals, and a large number of stakeholders. In spite of its importance, data from arable fields are not organized in such a way that they can be optimally used. Most systems are stand-alone, hindering easy access and (re)use of data. A grower goes to one consultancy firm, for example specialized in soil fertility for supplemental nitrogen dressing, and to another to be advised on irrigation. If these data elements were combined, the risk of leaching could be included in the advice, or even translated to a decision support rule. Clearly a common understanding of terminology and sharing of information is needed. In addition, forms required by the government and by the processing industry frequently overlap. So far this overlap has not been properly identified and exploited. Finally, we should restrict ourselves to farming only. The supply chain as a whole requires unambiguous terminology to achieve the required level of transparency. Lack of meaningful communication in general hinders new developments in terms of process optimisation, research and knowledge transfer.

With the appearance of databases in agricultural applications, consistent and coherent definitions of concepts have become essential for the agricultural community. After all, transferring data from one domain to another and merging data originating from different sources is hindered when concepts are defined in different ways. In the past, several attempts have been made to more or less formally organize the information related to arable farming. On the one hand, thesauri have been developed for structuring textual information, in particular for libraries. In agriculture, Agrovoc (FAO 2006) is a comprehensive and widely used terminological system. On the other hand, data models have been developed for use in databases. These data models were usually linked to specific software applications, and therefore not particularly useful for information sharing across different sources, systems and stakeholders.

This historical development can be seen as a prelude to ontologies. Whereas thesauri are too informal for computational processing and unambiguous interpretation, data models are too application-specific. Moreover, during the last decade both the penetration of the World Wide Web and advances in knowledge representation (object orientation, logics) have set the stage for a completely new approach to sharing and formalizing information. The term Semantic Web was coined to emphasize the two aspects of information sharing: semantics and connectivity. Nowadays, W3C (the organisation behind the semantic web) has set a number of standards for representing information and knowledge (XML, RDF/S, OWL) and for applying web services (World Wide Web Consortium 2006).

With the emergence of knowledge representation languages, the focus has shifted towards the definition of concepts, independent from particular software systems. 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.

At present, to our best knowledge no mature ontologies exist within the arable farming and processing domain. This holds for potato, but is equally true for any arable crop. Probably, modelling this domain is not so interesting for the academic world, whereas ontological modelling is too new to receive much industrial attention. The closest effort is probably the ontological version of the Agrovoc thesaurus that FAO is working on presently (FAO 2006). However, Agrovoc does not supply the detailed data elements we consider to be required in modern farming and processing business.

Second, work being done on ontologies in bioinformatics is progressing rapidly, the Gene Ontology being the most impressive example. However, these ontologies deal with internal biological processes rather than macroscopic production data. At this level a third area of ontological research is relevant here although not solving our problem, viz., the area of geo-information. For connecting GIS-applications and sharing GIS-data geometry-related ontologies are being developed (Seamless 2006). This work touches upon the ontology proposed here in the sense that it supports geometrical description of parcels.

To overcome the disadvantages of current practices in data mining, benchmarking and self learning systems and to benefit from the data collected anywhere we believe that a more structured approach is needed. First we need to compile and map existing data models (either loosely described or formally defined), analyse them for their mutual coherence and consistency. Next, we have to propose a unified model that relates all different entities and attributes, preferably in terms of an ontology as defined by W3C. The advantage is that all possible stakeholders will be able to understand the data expressed by this ontology and that software applications such as DSS's can process them automatically. It will also allow the application of advanced data mining techniques that may help to uncover previously unknown correlations.

In this paper we aim to provide an overview of existing data sources and data items. We describe the first version of an ontology for arable farming. In particular we focus on data pertaining to potato production for processing. Growing potato (in The Netherlands) for production of crisps or chips is one of the most profitable arable crops in the country. The area approaches 50,000 ha per annum.

The structure of this paper is as follows. First, we define our approach in terms of material and methods. We describe several areas in arable farming and the data elements involved in these areas. In Chapter "An Ontology for Data Management in Arable Farming" we show how these different data structures can be represented in a single ontology. We discuss the potential benefits of a shared domain model in "Discussion" and finally conclude in the "Conclusion."

Materials and Methods

Definitions and Structure

In this section we will provide an overview of the data items used by different sources of information in potato production and processing gathered in The Netherlands. The diversity of formats and models underlying these data will be used as raw material for setting up a unified model (ontology). Before presenting the information items we collected, we give a short terminological clarification. We have experienced that this interdisciplinary type of research sometimes induces unnecessary confusion with respect to the following notions.

- A *Model* is a description of some aspects of a real system (for example yield, shape, physical characteristics, information processing). By definition a model gives a simplified representation of reality. Models may be implemented in software, but not necessarily.
- A *Decision Support System (DSS)* is a software application that combines measurements with a model or database and allows user friendly interaction. The outcome of a DSS is a quantitative or qualitative analysis that assists a farmer in taking a decision (e.g., on treatment frequency, dose to be applied)
- An *Ontology* is a controlled vocabulary of a domain that describes concepts, attributes and the relations between them in a formal way using a standardized knowledge representation language.
- A *Self-Learning System* is an automated system that automatically improves its own knowledge or model base by systematically monitoring and analysing its performance, detecting improvements and deficiencies in the knowledge models and databases.

To obtain the data models needed for our study we conducted several interviews with growers, representatives from processing and consulting firms, representatives from inspection authorities and civil servants enforcing governmental policies. They supplied us with documents in which the information requirements, data input and data output were shown. Moreover, we studied the forms for on-line registration systems. Increasingly growers have to provide information to processors, governmental and certifying bodies through online forms.

Table 1 shows a list of stakeholders involved in arable farming, their goals and which means they have at their disposal to achieve them. It is obvious that the means always include the generation of and access to data.

As an example of the data flows between different stakeholders Fig. 2 schematically represents the flow of data within and around a potato field. The figure shows which mandatory data need to be collected by the grower as a licence to produce. The open arrows

Table 1 Actors, goals and means

Actor	Goal	Means
Farmer/grower	Maximising yield	Growing advice (DSS)
	Minimising liability	Application of inputs
		Storage advice
		Benchmarking
Processor/procurement	Risk control	Traceability of production data on farms
	Minimising liability	Benchmarking
	Maximising processing yield	Maximising recovery
Government	Check on regulations	Access to relevant production data regarding regulations
Certifying body	Check on certification conditions	Access to relevant production data regarding certification schemes
Consultant	Provide growing advice (DSS)	Access to relevant production data and production results
	Improve quality of knowledge	Data analysis capability
Trusted third parties	Facilitate stakeholders	Data warehouse
	Access management	Data analysis capability
		Application management

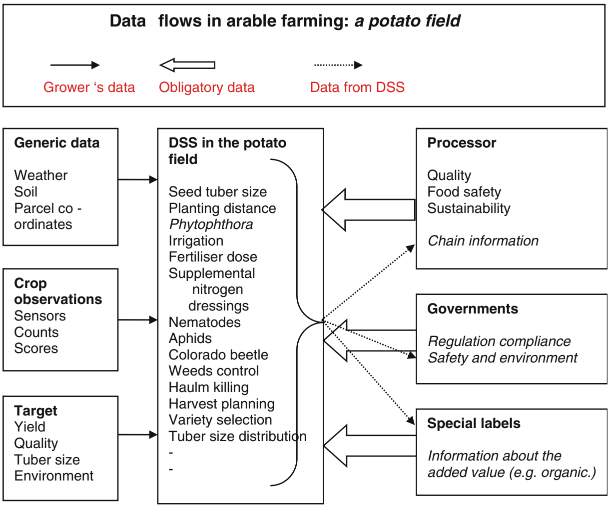

Fig. 2 Schematic representation of the flow of data within and around a field. The *open arrows pointing to the left* represent the requirements imposed by various stakeholders

pointing to the left indicate the conditions exerted by processors, governments and certifying bodies that growers have to fulfil. They grant the grower the licence to produce for a specific purpose provided that he or she complies with their requirements. Compliance is verified by giving access to the grower's data base that – among other – consists of the data contained in DSS's. DSS's may be of a strategic nature (e.g. growing seed or starch potato), tactical (pre-planting decisions such as which variety to grow) or operational (after planting). All DSS's need generic data, crop observations and produce targets supplied by the grower as shown in the three frames at the left side of Fig. 2.

In the following subsections we provide an extensive description of data sources. Based on these data sources, in the next section we propose a unified model in the form of an ontology for data management in arable farming.

Description of Mandatory Data

The processing industry is a major demander of data from a field. Growers have to complete paper forms or they may provide data online. This can be done through farm management systems, as provided by various suppliers in Western Europe. These companies in The Netherlands apply the standard system of registration called Edi-teelt (A standard for Electronic Data Interchange in the arable farming sector). Edi-teelt improves the quality of information exchange and reduces the costs of collecting it. Growers can exchange information with co-operatives and in future probably with governmental bodies.

The Dutch Society of Potato Processing Industry (VAVI) has constructed a certificate based on EUREPGAP that allows an HACCP-methods (Hazard Analysis and Critical Control Points) to be applied. Table 2 includes information on the risk of foreign bodies from machines, shotgun pellets, faeces and on chemical contamination such as from biocides, fertilizer, fuel and sewage sludge. Regulations ensure that no contamination with genetically modified (GM) crops and other mixes or exchanges take place.

Table 2 shows the information required by the VVA certificate that processors of potatoes have to supply with crisps and chips upon delivery. These firms use the forms that

Table 2 Food safety requirements of the potato processing industry (VVA certificate VAVI (2004)) GMC= Genetically modified crop

Subject	Risk	Measures by grower
General		
Machinery	Contamination by foreign bodies (use not in GMC)	Clean machinery, food grade oil, no contact with GMCs
Traceability	Mixing or switching crops	Registration of parcel, batch in store, keep documents for 2 years
Accidents	Foreign bodies (quarantined and GM crops, green tubers)	No mixing of batches (control, tuber removal from stores, machines and mandatory registration of accidents and corrective measures
Cultivation		
Parcel	Contamination (e.g. sewage sludge, glass, oil, metals, biocides, GMC, etc.)	Use of uncontaminated fields not near GMC, mandatory removal
Seed potatoes	Quarantined and GM crops	Certified seed of non GMCs
Fertilization	Excessive nitrate, biocides, PCBs, heavy metal.	Recommended doses following soil analysis, adjusted spreader, approved organics only, no sewage sludge
Crop protection	Overdose, record keeping, follow recommendations	Apply when necessary, correct dose with adjusted sprayer, observe waiting time before harvest
Hunting	Shotgun pellets in product	Shot not allowed to enter product
Water	Chemical and bacteriological contamination	Good water quality , in doubt consult analysis report
Harvest and warehousing		
Harvest, transport, store	Contamination by foreign bodies Mixing and exchanging	Sailcloth when raining, clean unbroken trailers and boxes, tuber removal between batches, no pets and vermin, no storage of tubers with chemicals, protect lamps,...
Lamps	Splinters	
Direct daylight	Green tubers	
More on storage (faeces of vermin, mercury from thermometers, sprout inhibitors,....)		
Ex-warehousing and delivery		
Foreign bodies	Contamination	Weed out at delivery
Safety period	Overdose of crop protection agents	Comply with safety periods

growers have completed to verify whether there is an increased food safety risk involved with a cargo of potato. They may want to use it for tracing and tracking purposes. If the form is indicative of a certain risk of contamination that specific lot is scrutinised and, e.g., sampled more intensively. The data in Table 3 have to be supplied when the potato lot is delivered. It deals with managerial and environmental information that determines the yield and quality of the crop and less with food safety as taken care of by the VVA certificate. The history of a parcel gives an indication of soil health and fertility. The seed quality (health and sprouting conditions) and seed rate (seed size and planting distances) are related to yield and tuber size as are management practices such as fertilization, irrigation and crop protection. Keeping track of this information and comparing growers allows processors to identify successful strategies that growers have and to source from areas and types of growers that best meet their requirements. Therefore they have to make observations in the potato production, leading to the data items presented in Table 3.

Table 3 Data requirement for potato producers for processing companies (compiled from forms growers have to complete at delivery of their ware potatoes)

Item	Observation
History of parcel	Last potato crop: when, variety
	Last year's crop
	Use of green manure
	Soil fumigation, when, active ingredient, dose (kg ha^{-1})
Equipment	Date testing of sprayer and spreader
	Licence to spray of applicant
	Type of irrigation equipment
Soil	Date soil was sampled last and soil type
	Percentage organic matter
	Percentage clay
	pH-KCl, Phosphate (Pw), Potassium (K-number), Mg
Planting and seed potatoes	Distance between rows and plants
	Seed tuber size
	Seed potato grower number (see seed certificate)
	Sprouts: no, small white sprouts or green pre-sprouted
	Seed treatment against black scurf (code, active ingredient, dose kg t^{-1})
Irrigation	Number of irrigations
	Number of mm per irrigation
	Type of water (deep well, surface, other)
Manure and or compost	Date of application and per application date:
	Minas code
	Type of manure
	Solid or liquid
	Dose (t ha^{-1})
	Mineral concentration (pure N, P_2O_5, K_2O)
Chemical fertilisers (as of autumn)	Date of application and per application date:
	Type of fertilizer
	Amount (kg ha^{-1})
	Pure N, P_2O_5, K_2O, Mg in fertilizer (%)
Foliar chemical treatments	Weed, insect, late blight control, haulm killing:
	type and date of treatment
	product and active ingredient
	dose (kg ha^{-1})
	applied by grower or contractor
Mechanical weed control	Date of hilling
Storage	(Expected) Harvest date
	Sprout suppressant (application dates, type, dose, temperature)

The assessment of product quality as carried out by the processing industry is shown in Table 4. The data in Table 4 are compiled from analyses of samples taken from loads collected at farms. Gross and net yield are determined and the sample is graded. Grades vary in destiny and value. The processing quality is higher when tubers have a higher dry matter concentration, when less of them float in brine and when they have lower concentrations of nitrate and reducing sugars. After cooking darkening (ACD) and fry colour (FC) both should be light and higher length–width ratios increase recovery, i.e., kg's of finished product per kg raw material used. Finally, a lower proportion of defects increases the value of a lot.

Table 4 Observations on tuber quality made by potato processing industries in samples of procured potato loads

Specification	Observation
Yield	Tuber mass (t ha^{-1})
	Tare (%)
Size grading	< 40, 40–55, 55–70, 70–90, > 90 mm
Processing quality	Specific gravity or dry matter concentration (%)
	Tuber nitrate concentration (%)
	Floating tubers in brine with SG 1.072 (18.5 % DMC (%)
	Concentration reducing sugars (specific products only)
	Fry colour (ranks on USDA chart)
	After cooking darkening (1–9 scale)
Defects (%)	Black spot and bruising, silver and scab (% per severity class)
	Rotting tubers, tubers with growth cracks
	Internal rust spot (% per severity class)
	Green tubers, hollow tubers, tubers with sprouts > 1.5 cm
	Spraing. sugar ends

The last set of mandatory information items a grower has to supply as 'a licence to produce' are the governmental requirements shown in Table 5. These data were taken from the e-counter at the website of The Netherlands Ministry of Agriculture. Growers can download forms or complete them online. The coordinates of each parcel are registered centrally and these coordinates (vectors) are stored centrally but also supplied to the grower to allow verification. Within each parcel growers may grow several crops and each year in May (the so-called May Count) they have to indicate which crops are grown where. Crops may be subjected to a quota system (starch potato and sugar beet), or a premium may be given for fallow or production of non-food, non-feed crops. The type of crop has repercussions for the mineral balance aimed at reduced risk of leaching because a threshold of 50 ppm nitrate in ground water holds. To protect foreign trade – especially of seed potatoes – any occurrence of quarantine (Q-) diseases has to be reported and temporal and spatial actions to mitigate their effect need to be reported. Also sometimes when a disaster takes place such as an outbreak of brownrot or inundation, specific information needs to be supplied to receive relieve funds. General Food law requirements of the government are taken care of by the VVA certificate (Table 2).

Description of Data Needed for Tactical Decisions

To assist them in making decisions, growers have access to several public data sets, either for free or by paying a fee. This includes environmental data such as weather and soil conditions, used for strategic decisions such as the suitability for one farming type or another. Growers also have access to data to base their tactical decisions on (e.g., what profitability to expect and where to grow which variety in which amount) and pre-planting decisions regarding nematode control. For these decisions two important datasets are available: KWIN (quantitative information, Dekkers 2002, Table 6) and the Catalogue of Varieties (Table 7).

Data for Economical Return Assessment

Growers calculate net margins of growing a potato crop by using KWIN data. All current prices of inputs such as land, labour, machinery and chemicals as well as services to improve

Table 5 Government requirements, compiled from the e-Counter of The Netherlands Ministry of Agriculture (http://www.LNV.nl)

Rule	Objective	Information supplied on
Field registration	Mineral bookkeeping	Field coordinates
(May Count)	Subsidies (MacSharry)	Crop per field
Authorization	Risk avoidance	E.g. GM crops
(International) trade	Phytosanitary measures	Location of quarantine diseases and pests (m)
Mineral balance	Reduce risk of N and P emission	Soil type (sand, loess, peat)
		Maximum mineral N, P and manure allowed (kg ha^{-1})
Pre-planting crop protection plan	Reduce risk of biocide emission	Soil borne diseases, pests and weeds
		Cyst nematode sampling
		Seed potatoes health and Quarantine disease
		Hygienic measures
		Rotation frequency
		Crop inspection (diseases, weeds, aphids,..)
		DSS: late blight, aphid counts, weed control
		Mechanical weed control (pre-sprouting, late ridging)
		Chemical treatments
		seed potatoes (*Rhizoctonia*, aphids)
		choice of chemicals (reckoning with environmental effects
		site specific application
		emission reduction (crop free zone, thrift poor nozzle,..)
		recording in a log book where was deviated from the plan and why
General Food Law (GFL)	Food safety	Biocide, type, dose and time of application (see VVA certificate)
Common Agricultural Policy (CAP)	Stimulate markets, income and environment	Fallow or non-feed, non-food crop
		Nature conservation measures (flora, fauna)
		Cross compliance
		Rural development measures
		Innovations
Relieve	Mitigate damage	Quarantine diseases (brown rot) (€ ha^{-1})
		Damage by fauna (€ ha^{-1})
		Inundation (€ ha^{-1})
Revenue service	Tax collection	Costs (€ ha^{-1})
		Benefits (€ ha^{-1})

management practices through decision support systems are given in tables. If a grower would record all acts, such as the time required and machinery used and all inputs such as amount of biocides, he is able to predict the profitability of his operation. KWIN allows the grower to transform the records needed for the certifying agent, the processor and governments into the financial repercussions. Table 6 shows the meta-data stored in the KWIN database. Under 'market' prices of market driven inputs and products are given. Prices of chemicals are mean sales values of those of the most marketed trade marks and active ingredients. Soils are sampled for the presence of nematodes (see also Table 8) and to assess the amount of minerals (N, P, K) prior to planting (see also Table 12) at a costs charged by the labs carrying

Table 6 Information available in KWIN: quantitative financial information for arable farming (compiled from Dekkers 2002)

Subject	Item
Market	Price of land (buying or leasing) € ha^{-1}
	Average yield t/ha per area (last 10 years)
	Farm sale price (€ t^{-1}) 2-monthly average last 5 years
	Seed potato prices delivered on farm
	free market varieties
	licensed varieties
Chemicals	Fertilisers (€ t^{-1})
	nitrogen, phosphorus, potassium, compounded fertilisers
	lime
	other fertilisers
	Biocides (€ kg^{-1})
	seed piece treatment (*Rhizoctonia solani*)
	soil fumigants (nematicides)
	herbicides, fungicides, insecticides, haulm desiccants
	sprout suppressants
Soil sampling	Price in € ha^{-1}
	potato cyst nematodes
	root knot nematodes and free living nematodes
	pre-planting mineral assessment
Storage	Grading (€ t^{-1})
	Storage (€ t^{-1} $month^{-1}$)
	Losses (%$month^{-1}$)
	Energy need ($kWht^{-1}$) heating and cooling
Machinery	Tractor € ha^{-1} or h^{-1}
contracting	Fertilizer and manure application equipment
leasing	Soil treatment (ploughing, harrowing, hilling)
	Planting, spraying, irrigating, harvesting, store loading and de-loading

Table 7 Variety related information in The Netherlands catalogue of potato varieties (Baarveld et al. 2003)

Variety property	Range
Maturity	9=very early, 3=very late
Foliar development	9=very good, 6=average
Skin colour	dy=dark yellow, br=reddish brown
Tuber shape	R=round, lo=long oval
Eye depth	9=very shallow, 3=very deep
Tuber size	9=very large, 4=very small
Dry matter concentration	9=very high, 4=very low
Cooking type	A=firm, D=very floury
Consumer quality	fresh, crisps, chips, salad, mashed, processing, starch
Resistance to:	R=resistant
Virus (X, Yn, LR)	9=very good resistance, 3=very susceptible
Leaf and tuber blight	
Common scab	
Internal bruising	
Wart (Fysio 1)	O=immune, V=susceptible
Nematodes	Resists (combinations) of Ro1,2 and 3 and Pa 2 and 3

Table 8 DSS Nemadecide data input and output for nematode control (Been and Schomaker 2004)

Input		Output (decision support rules)
Soil	pH	Yield (kg ha $^{-1}$ and € ha $^{-1}$)
	Soil type (sand, peat clay)	Pop densities Pi and Pf (counts g $^{-1}$)
	% silt and % organic matter	Risk analysis (% chance at % damage)
	Bulk density (g ml^{-1} from above)	Rate of return on investment in nematode control (€ €$^{-1}$)
		Optimal rotation (crops and frequency)
		Optimal next sampling date
Nematode	Species	Break down of field in smaller polygons (depending on sampling density) suitable for Plant Health Service Declaration
	Pathotype (e.g. ro1, pa2,...)	
	Cysts (cc $^{-1}$)	
	Eggs (cc $^{-1}$)	
	Living larvae (cc $^{-1}$)	
Field	Vectors of polygon of	
	parcel (x,y,z (m))	
	potato plot (x,y,z (m))	
	Rotation (potato frequency)	
	History (nematode densities previous samplings (n cc^{-1})	
Sampling method	Company	
	Nak-Agro	
	Groene Vlieg	
	BLGG[a]	
	Method	
	e.g. AMI 50	
Nematicides	All types	
	KWIN information	

[a] Soil analysis company (Bedrijfslab voor Grond en Gewas).

out these services for growers. Prices of machinery – with and without driver – such as tractor, plough, harrow, planter, sprayer, harvester and store loading equipment are given per hour and or per hectare.

Data for Variety Selection

The *Netherlands Catalogue of Potato Varieties* (Baarveld et al. 2003) shows the most important characteristics of potato varieties for a grower to base the decision on which variety to grow. Once grown the information in Table 7 is part of the parcel database. The decision which variety to grow is based on market demand for tuber characteristics (size, shape, colour and cooking type) and on managerial decisions to be made concerning crop health and crop protection (Table 7).

Data for Potato Cyst Nematode Control

The presence of potato cyst nematodes is a reason for concern. In seed potato areas it is a quarantine pest and is absent and in other areas the presence and damage can be controlled using a DSS *Nemadecide* (Been and Schomaker 2004). The input for this system is described in Table 8. Input includes soil characteristics such as pH and bulk density, sampling method, encountered nematode species, pathotype and population density and its spatial distribution.

Table 9 Model input requirements and model output used for crop monitoring and yield forecasting (DSS Tipstar, Achten et al. 2005)

Type	Observation
Crop data	
Crucial dates	Days of planting, emergence and of (periodic) harvest
Sampling	Number of tubers
Meteorological data	
Temperature	Daily maximum and minimum temperatures (°C)
Moisture	Precipitation (mm day^{-1})
Solar radiation	Average daily wind speed (m sec^{-1})
	Relative humidity or vapour pressure deficit (kPa)
	Global radiation (W m^{-2} day^{-1})
Soil data	
Water retention per soil layer of 10 cm based on pedotransfer functions	Particle size distribution
	proportion clay silt and sand
	texture class: very fine, fine medium fine, medium or coarse
	Organic matter concentration (g g^{-1})
	Bulk density (g/cm^3)
	Depth of rooting zone, horizon, water table drainage (cm)
Soil fertility related data	pH_{KCl}
	Cation exchange capacity (meq/100 g)
	Pre-planting organic matter concentration (g g^{-1})
	Pre-planting mineral nitrogen concentrations (g g^{-1})
Model output	
Organs (leaves stems and tubers)	Leaf area (cm^{-2})
	Proportion light interception (%)
	Rooting depth (cm)
	Daily dry matter yield (g m^{-2})
	Dry matter concentration (%)
	Nitrogen concentration (mg kg^{-1})
	Tuber specific gravity (g g^{-1})
	Tuber protein and starch concentrations (%)
Soil daily values	Moisture content per layer (mm cm^{-2})
	Mineral nitrogen concentration including residual N (mg kg^{-1})
	Nitrogen volatilization (mg m^{-2})
	Nitrogen leaching (mg m^{-2})
	Water drainage

Output are scenarios of future population densities, yields and rate of return of control measures depending on nematicide application, choice of variety and frequency of potato cropping. The model also recommends when the soil should be sampled again.

Description of Data Needed for Operational Decisions

Once the crop is planted many operational decisions need to be made regarding control of weeds, pests and diseases, mineral and water supply achieved yield to date and harvest planning.

Table 10 Input and output for weed and haulm killing according to the DSS minimum lethal herbicide dose (MLHD, Kempenaar and Van den Boogaard 2004) and Gewis (Opticrop 2005)

Input	Output (decision support rule)	
Weed killing		
Weed species	Sensitivity to herbicide	Herbicide dose (kg ha $^{-1}$)
	not sensitive (−)	
	moderately sensitive (+)	
	sensitive (++)	
	very sensitive (+++)	
Weed stage	Seedling (cotyledons only)	
	2 leaves	
	4 leaves	
	6 leaves	
	> 6 leaves	
Post treatment sensor (PS1)	Proportion absorbance at 820 nm	
	indicative of photosynthetic activity	
Herbicide	Choice e.g. Sencor	
Haulmkilling		
Sensors	Cropscan (crop reflection sensor)	Herbicide dose (kg ha $^{-1}$)
	WDVI (Weighted Difference Vegetation Index)	
	PS1 (crop absorption sensor)	
	Proportion absorbance at 820 nm	
Herbicide		
Gewis (not an acronym, pun for Dutch words "crop and sure")		
Environment	Radiation (W m^{-2} day^{-1})	0–2 scale of Wax layer
	Rain (mm day^{-1})	Hydration
	Wind (m sec^{-1})	Leaf wetness
	Soil temperature (°C)	Photosynthesis
	Crop temperature (°C)	Uptake of biocide
	Soil surface wetness (+/−)	Time of application
Biocide	Choice	Relative dose

Data for Crop Growth Assessment and Yield Forecasting

LINTUL-POTATO (Kooman et al. 1996) is a summary model that computes crop development given temperature. Sprout growth rate prior to emergence is temperature driven and so is the leaf extension rate after emergence and leaf duration before senescence. Variety, temperature and day length determine the moment of tuber initiation and crop earliness as earlier crops distribute more dry matter to the tubers at the detriment of the foliage that necessarily dies earlier. Dry matter accumulation is determined by the amount of solar radiation and the radiation use efficiency that is negatively influenced by high temperatures (increased respiration) and low availability of nitrogen and water. The two latter factors also have an effect on leaf longevity hence interception of light by the canopy. The model use daily values of real time weather to monitor current growth and uses long term historical weather data to explore future growth and risks of extremes.

An example of the use of the LINTUL-Potato model is TIPSTAR (Table 9, (Achten et al. 2005)), a DSS in starch potato production that enables crop monitoring and yield forecasting. Subsequently TIPSTAR calculates the requirements of nitrogen and water at any moment during growth. The system also forecasts nitrogen leaching through the amount of water drained and its nitrate concentration.

A yield determining factor is the water holding capacity of the soil. In almost all years evapotranspiration exceeds precipitation. Consequently part of the water has to be removed from the soil stock. This stock depends on the depth of the rooting zone and the water concentration per layer of the zone on its turn dependent on the particle size distribution (pedotransfer functions, (Makkink 1957, Vereecken et al. 1989 and Wösten et al. 1995)).

Data for Weed Control

Several systems support herbicide dose application (Table 10). A first method is *MLHD* (Mimimul Lethal Herbicide Dose, Kempenaar and Van den Boogaard 2004), to be applied both for weed and haulm control. This method applies the minimal amount of herbicide needed to effectively kill the weeds, while causing as little damage to the crop as possible. Using this method may lead to a saving of up to 75% of active ingredient.

The approach is based on a quantitative model that has shown that weed susceptibility to the herbicide depends on species and age (cotyledons and leaf stage). From these two parameters as input, an herbicide dose is recommended. Two days after application the grower tests the efficacy of the low dose with the aid of a PS1 meter. The PS1 meter shows whether the photosynthesis mechanism is still functioning or not and indicates the need for additional spraying. For haulm killing the recommended dose depends on the size of the canopy as sensed with a crop scan at two wavelengths. As with MLHD an early warning for its efficacy is shown within two days.

The DSS *Gewis* (Opticrop 2005) uses as input environmental data such as temperature and soil wetness to determine foliar properties that determine the uptake of an active ingredient hence the best time to apply and its relative dose.

Data for Late Blight Control

Two major companies exist in The Netherlands that offer decision support systems to advise on control of late blight caused by *Phytophthora infestans*. This disease is responsible for the bulk of chemicals in potato production and adds 20% to the production costs. Dry and hot spells reduce the need to apply whereas growers most often now apply fungicides on a regular (weekly) interval. The DSSs are Plant Plus of the Dacom company (Raatjes et al. 2004) and Prophy of the Opticrop company (http://www.opticrop.nl) Table 11 show the data requirements and output of the combined two systems. The duration of leaf wetness following dew, precipitation and overhead irrigation as further influenced by temperature, relative humidity and wind speed are the main determinant factors for successful spore germination. Late blight development consists of germination and penetration of spores, growth of the micro-organism inside the leaves and formation of new spores. The expected severity of the disease depends on (a) disease pressure i.e., proximity of the field to the nearest outbreak, (b) temperature and humidity (rainfall and dew) and (c) variety. The expected degree of control by fungicides depends on the type of fungicide – systemic or contact – and the effective coverage of leaves with the chemical. This depends on leaf growth rate (newly formed leaves are not covered) and run off of the chemical due to precipitation. Outputs of the DSS are pressure, risk of an outbreak, advice to spray and when, type of fungicide and dose.

Table 11 Input and output for late blight control Plant Plus (Raatjes et al. 2004) and Prophy (Opticrop Company in The Netherlands, http://www.opticrop.nl/) and irrigation planner (Opticrop Company in The Netherlands, http://www.opticrop.nl/)

Input		Output (decision support rules)
Late blight[a]		
Hourly weather data (last few days and forecast)	Temperature (°C)	Risk of outbreak (0–100 scale)
	Precipitation (mm)	Disease pressure (average risk last seven days)
	Relative humidity	Duration of the positive effect of the last treatment (days)
	Leaf wetness (+/−)	
	Wind speed (s^{-1})	
	Wind direction	
	Radiation (W m^{-2})	
Crop	Variety (and its susceptibility)	Advise to control: yes, maybe, not
	Emergence date	Type of fungicide
	Sprinkler irrigation (data and mm)	Dose
	Last fungicide application (type and dose)	
	Leaf growth rate	
	scale 1–5, 3=average)	
	Ground cover	
	scale 1–7 between 0 and 100 %	
	scale 8–10 lodging 0–100 %	
	Or stem growth rate	
	scale 0–10 between 0 and 20 cm/week	
	Crop vigor	
	<, =, > vigorous than control crop	
	Crop stage	
	Scale 1–10 from sprouting to maturity	
	Vicinity of nearest outbreak	
	scale 1–10 between 10 miles and in field	
Irrigation		
Weather	Evapotranspiration rate (mm day^{-1})	Soil moisture condition (pF)
Soil	Soil types (several layers possible) and corresponding water holding capacity (mm mm^{-1})	Irrigation advise between pF 2.7 (start of drought) and pF 3.2 (severe damage level): either or not
	Water table (cm) for capillary rise MUST model (De Laat 1980)	
Crop	Date of emergence	Timing of irrigation amount (mm)
	Last date and amount of irrigation	
	Rooting depth (cm)	
	Ground cover (%)	

[a] The DSS's differ in input and output, here combination of the two is shown.

Data for Irrigation Management

An irrigation planner (second part Table 11) such as developed by Opticrop (http://www.opticrop.nl) uses similar meteorological data as the late blight planner and those shown for mcrop assessment and yield forecasting in Table 8. Capillary rise is calculated and further input is crop emergence, rooting depth – for water availability and proportion ground cover which is proportionate to the evapotranspiration (ETP) rate. With a forecast of ETP for the next few days the DSS recommends timing and amount of water to be applied.

Data for Fertilization Management

The most widely used DSS (Stikstof^{plus}, Table 12) aimed at improving timing and rate of organic and chemical fertiliser has been developed and is exploited by the BLGG company (Bedrijfslab voor Grond en Gewas) at Oosterbeek, The Netherlands, (http://www.blgg.nl). The DSS determines the pre-planting amounts to be applied to the soil of macronutrients N, P and K, meso-nutrients S, Mg and Mn, the timing and amount of supplemental nitrogen to be applied after planting and foliar dressings of macro-, meso- and micronutrients. The soil of the parcels or plant parts are sampled following statistical procedures and subjected to chemical analysis to determine the concentration of the various minerals. Nitrogen

Table 12 Soil and crop properties and recommended doses of minerals from the DSS Stikstof^{plus}, BLGG company in the Netherlands (http://www.blgg.nl)

Soil input		Output (decision support rules)
Pre-planting nitrogen	Soil stock	N release(kg ha−1)
	Depth of sampled layer	from organic manure
	Pre-planting N-ammonia (kg ha^{-1})	from green manure
	Pre-planting N-nitrate (kg ha^{-1})	from the previous crop
	Organic nitrogen (mg kg^{-1})	from soil organic matter
	Yield expectation (t ha^{-1})	Recommended time and dose
	Organic manure (t ha^{-1})	(kg ha−1)
	Green manure (type, when)	
	Previous crop	
	Variety and/or crop type (seed, early, main)	
Supplemental N-dressings	Soil stock (kg ha^{-1})	Recommended time and dose (kg ha−1)
	Or:	
	Petiole nitrate concentration (mg kg^{-1})	
	Or:	
	Canopy mass (t ha^{-1})	
	Canopy nitrogen concentration (%)	
	Or:	
	Crop reflectance	
Mineral elements	Phosphorus (P), Potassium (K), Sulphur, Magnesium (Mg), Sodium (Na) (mg kg^{-1})	Recommended dose prior to planting (kg ha^{-1})
	P-AL (mg P_2O_5 per 100 g	
	C/N	
	Sulphur, Copper, Borium, Zink (μg kg^{-1})	
	Desirable range per element	
Physical properties	PH	Amendments of lime, organic matter (t ha^{-1})
	Organic matter	
	Cation exchange capacity CEC (mmol kg^{-1})	
	Granular composition[a]	
	Electric Conductivity (EC) S m^{-1})[a]	
Crop input during growth	N, P, K, Mg, Ca, S, Cl, Fe, Mn, B, Cu, Zn (mg or μg kg^{-1})	Foliar spray dosage (kg ha^{-1})

[a] Can be done but is not done routinely.

recommended doses depend on expected crop uptake – a grower has to input expected yield – and nitrogen supplied by the soil. The latter is from the amount of pre-planting nitrogen (nitrate and ammonium), from mineralization of soil organic matter (soil organic nitrogen and C/N ratio are input) and information regarding the previous crop and the use of organic and green manure. When multiple dressings are applied, only part of total calculated dose is applied at planting and during the season either soil or crop is assessed through sampling of minerals or reflectance (Booij and Uenk 2004) whereupon a second dressing is calculated and recommended. Many soil labs also determine physical soil properties and where deficient recommend remediation by applying lime to increase pH or organic matter to improve soil fertility and water holding capacity.

An Ontology for Data Management in Arable Farming

The previous section has presented an overview of the data items defined by a diversity of information sources and data users for arable farming. We have taken the situation in The Netherlands as an example, but probably many concepts and relations are similar in other countries. Careful inspection of these items learns that all sources only refer to a limited number of central concepts, relating to real world entities: farm, parcel, crop, soil, etc. However, different parties have different views on these terms and this has caused a multitude of interpretations. As stated before, a single information model (ontology) will eliminate confusions and clarify unclear notions. Even more interestingly, an ontology in this area will facilitate unambiguous communication between computers and support automation of many processes.

Ontologies come in different shapes and tastes, depending on the level of formalisation and the knowledge representation used. Here we have decided to employ OWL (Web Ontology Language) for expressing the data items presented above. OWL is the most advanced standard for the Semantic Web, introducing features from description logics into frame-based (object-oriented) representations. However, for explaining the main structure of our ontology we need not go into all technical details of OWL. In the following subsections we will restrict ourselves to explaining the main *concepts* (classes) and their *properties* (relations). At this moment we only mention a few *instances* for illustration purposes. We also describe a number of *restrictions* that constrain the values that properties of subclasses can have, in comparison to their super classes.

In the following section we describe the organisation of concepts and relations in arable farming, with the potato crop as a particular example. We have employed Protégé (2006) as a modelling tool for the analysis of the terms mentioned in the previous sections. Protégé provides a user friendly environment for defining concepts, relations, cardinalities, etc. It supports a number of verification checks and generates the formal model description in terms of RDF/S and OWL. Some of the graphical representations were generated by the visualisation component called Jambalaya. Note that not all details of the ontology are shown for clarity reasons.

CROP ONTOLOGY – An Ontology for Arable Farming

Figure 3 shows an overview of the basic concepts we identified in this domain. A *Farm* has multiple parcels. A specific *Parcel* has a layout, soil, meteorological profiles and treatments. The layout of a parcel describes its coordinates, the distance between plants

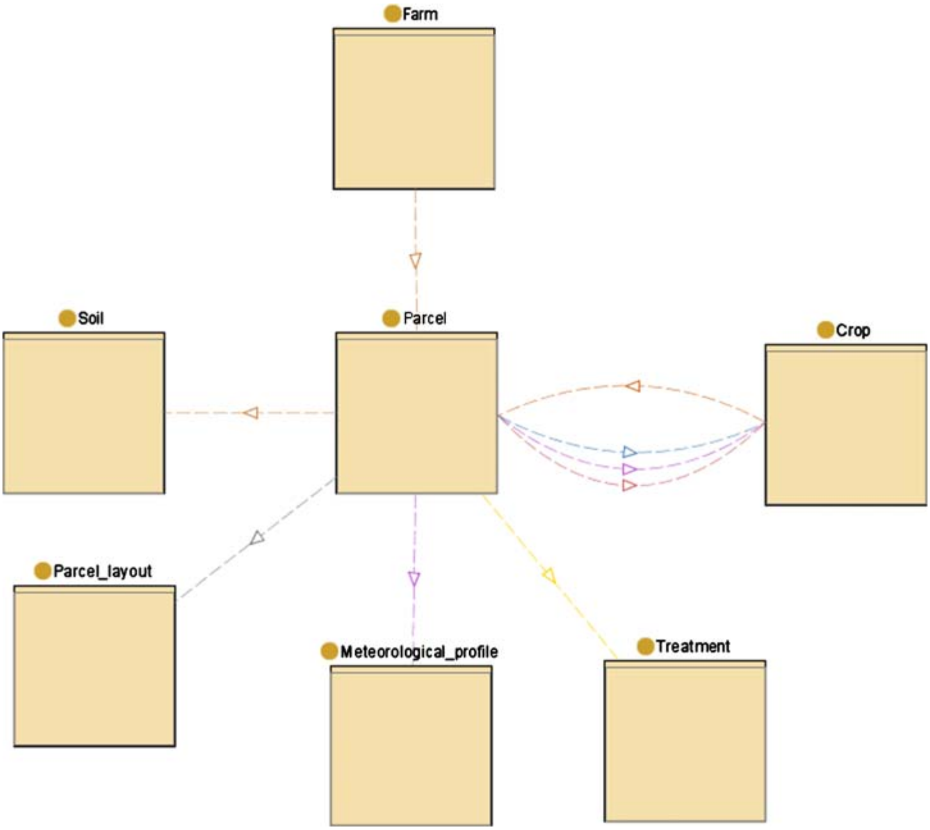

Fig. 3 Basic concepts in the ontology for arable farming

and the distance between rows. The other concepts will be described below. The diagram shows that a parcel has a complex relation with the concept *Crop*. A parcel may have a present crop, a last year's crop and a last crop. Moreover, we concluded that a crop should have an inverse relation with a parcel, since it is necessary to know for a crop from which parcel it originates.

The complete set of attributes of *Parcel* is also shown in Fig. 4. Note that in the ontology the *cardinalities* are indicated, as shown in Fig. 4. These cardinalities define whether an attribute can have only one value or multiple values. For example, a parcel only has one layout, but may have undergone several treatments. This figure also shows that the value of each attribute must be of a specific type. Obviously a treatment can only be of class *Treatment*, which in turn has several subclasses. In particular we have defined *Irrigation*, *Fertilisation*, *Foliar_chemical_treatment*, *Mechanical_weed_control* as subclasses of Treatment. These subclasses share the fact that they are applied by someone, at some date, using some equipment.

The concept Parcel also has a relation to *Soil*. It is important to note that an instance of Parcel, i.e., a specific parcel existing somewhere, contains an instance of Soil, not just a type of soil. Such a particular occurrence of soil contains a certain amount K, P, Mg, has a pH-level, all recorded at a certain date (see Fig. 5). The concept Soil has a number of

Fig. 4 The concept Parcel and its attributes

```
Parcel
  -   layout (single Parcel_layout)
  -   soil (single Soil)
  -   metereological_profile (multiple Metereological_profile)
  -   crop (single Crop)
  -   last_crop (single Crop)
  -   last_years_crop (single Crop)
  -   treatment (multiple Treatment)
```

subtypes (as for example *Loess*, *Peat* or *Sand*). So, in practice we can for instance define an instance of Loess with pH=7.

A *Crop* is of a certain species, for example potato, cabbage, wheat, etc. This is expressed by subclasses of the concept Crop (Fig. 6). Whereas the variety tells which properties a crop species has in general, or at least is expected to have, instance of the concepts Crop itself points to a specific crop growing at some parcel and harvested sometime. At the class level, maturity_timing tells what the expected duration of reaching maturity for a variety is. For an instance the actually obtained maturity_timing can be recorded. Note that in this case the expected value serves as a *default* value for instances of Crop. OWL presently cannot formally handle defaults as they conflict with the underlying logical representation. However, for our present objective we choose to use the restriction construct of OWL to represent this. Only if formal reasoners are used this may give rise to unexpected results.

A crop growing on a parcel is characterised by, e.g., its foliar_development and organ description. A Crop can have multiple defects, e.g., black spot, nematodes, scab, hollow tubers, etc., and a specific variety may be resistant to a number of defects. A crop further has size_grading and tare as its properties. Its quality for processing is described by the concept *Processing_quality*, containing relations such as after_cooking_darkening, dry_-matter_concentration, fry-colour, specific_gravity, and so on. Note that this may be different from the expected processing quality as specified for the variety. In the same way, the *Consumer*_quality can be described.

In our present ontology of arable farming we focus on *Potato* as a subclass of Crop (see Fig. 7) . In addition to all properties that are defined for a crop, we have identified several attributes specific for potato. For example, it is of a certain cooking_type, and may be stored using sprout_suppressing. The class *Sprout_suppressing* could have been modelled as a subclass of Treatment, but we decide not to do so. Treatments are applied to parcels by a specific applier; this is not the case for sprout suppressing.

A complication arises with modelling potato properties (Fig. 8) such as eye-depth, skin_colour, tuber_shape and tuber_size. We observe that in practice these terms are sometimes considered as a property of a crop and sometimes of an individual tuber. We have solved this ambiguity by proposing the property *average_tuber* for potato crop, summarizing properties that actually specify a single tuber. Note that the same issue arises for properties such as dry_matter, but as this is not normally assigned to a single tuber we have decided to move this to the property processing_quality of the crop as a whole.

Fig. 5 The concept Soil and its attributes

```
Soil
  -   K-contents (single Concentration)
  -   Mg-contents (single Concentration)
  -   P-contents (single Concentration)
  -   pH (single pH)
  -   clay_contents  (single Concentration)
  -   organic_matter_contents (single Concentration)
  -   date_last_sampled
```

Fig. 6 The concept Crop and its attributes

```
Crop
   -  parcel (single Parcel)
   -  maturity_timing (single Scale_3_to_9)
   -  resistant_to (multiple Crop_defect)
   -  defect (multiple Crop_defect)
   -  foliar_development (single Foliar_specs)
   -  organ (single Crop_organ)
   -  harvest date (single Date)
   -  size_grading (single Size_distribution)
   -  tare (single Tare)
   -  processing_quality (single Processing_quality)
   -  consumer_quality (single Consumer_quality)
```

However, this decision is somewhat arbitrary and future use of the ontology may lead to different designs in this respect, depending on the application context.

Other concepts currently defined in the ontology are for example Equipment, Fertilizer, Metereological_profile.

Presently, the ontology on processing potato production covers roughly the items mentioned in the previous section. We see a number of ways to further extend the quality and coverage of the ontology:

- Mapping the concepts in this ontology to those defined in other domain models will help to find missing concepts and relations. Agrovoc (FAO 2006) can be used for this purpose.
- We intend to employ ROCK (Koenderink et al. 2005) as a method to elicit additional concepts and to refine the existing conceptualisation. ROCK stimulates experts to explicate statements on the domain in an associative way.
- Orchestrated feedback on the ontology by experts in arable farming and IT-specialists in this field. The optimal way of obtaining useful comments and extensions is yet to be established.

Discussion

What can a fully developed ontology mean for arable farming? Presently we see a rapid proliferation of ontologies in many different domains, as for example in libraries, museums, research (in particular in the context of genomics), etc. The benefits of having adequately defined ontologies can be attributed to both enhancing human interpretation and supporting machine processing.

A widely accepted standard for modelling the domain of arable farming helps to improve communication between consultants, researchers, farmers, government agents because it provides a shared vocabulary. A simplified version of the ontology, represented in a single overview (the ‘reference map of arable farming’) would already pay-off. It also would help users to interpret the use of software systems, such as decision support systems. For *software developers* on the other hand, an agreed-on ontology will provide a common basis for designing the internal architecture of a system, and even more importantly a consistent

Fig. 7 The concept Potato, subclass of Crop

```
Potato
   -  average_tuber (single Tuber)
   -  sprout_suppressing (single Sprout_Suppressing)
   -  cooking_type (single String)
```

Fig. 8 The concept Potato_tuber

```
Potato_tuber
   -   eye-depth (single Eye-depth)
   -   skin_colour (dark_yellow, reddish_brown)
   -   tuber_shape (single Tuber_shape)
   -   tuber_size (single Tuber_size)
```

framework for data exchange. Moreover, grounding a user interface design on a shared ontology improves recognition among different users. This would require a shared effort of software developers in the arable farming domain, to agree on a standard version of the ontology. The benefits, improved compatibility of information, would pay-off in terms of reduced design and operation costs, and improved quality of the available software.

However, the major effect of using the ontology of arable farming will be realized when it is directly integrated with a software system. First, intelligent search engines on the web are increasingly dependent on ontologies for finding *all relevant* information one is looking for. This replaces simple string-based searching. The ontology is using RDF/S (the layer below OWL) to point to concepts and relations between concepts. This implies that a world-wide unique reference (URI) is used to identify these concepts and relations. Search engines can locate this information for retrieving documents.

Second, the ontology can be used directly as a database using tools like Jena (Jena 2006), rather than employing the traditional relational database schema. This means that both the data (instances) and database schema (concepts and relations) are represented explicitly in the database, resulting in a highly flexible solution.

Finally, the full potential of ontologies comes available once they are integrated with web services (see e.g. (Korotkiy and Top 2006)), resulting in so-called *Knowledge Utility Services*. The ontology is extended with a number of services, effectively providing information on the domain. For example, in arable farming useful services could be to provide the characteristics of a certain variety of potato, to predict yield given certain conditions or to provide alternative varieties given changing environmental requirements. In fact, decision support systems will typically become available as web service in the next generation of the web.

Moving away from the technical consequences, the effects on society and farming in particular will also be manifold.

– Ontologies can help to minimize the administrative load. Information needed by governmental agencies can be streamlined with information flows within the supply chain.
– Self-learning systems as defined earlier in this paper can build on a shared ontology to integrate observations from different systems, and from production and processing sites world-wide.
– Research on arable crops and their production can profit by sharing results from experimental and theoretical science. The term e-science is already frequently used in this context. Moreover, ontological techniques are being developed to automatically extract knowledge from textual documents.
– Existing providers of DSS solutions can already employ the above ontology to obtain data from decentralized data repositories.
– 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).

When using data in a decentralized way to facilitate benchmarking, automatic data analysis and learning, data ownership and data security can become important restrictions. In the realm of data warehousing, these issues have been discussed, and adequate solutions are proposed in *e.g.*, Kimball and Ross (2002) and Inmon (2005). A number of operational chain information systems already use these solutions.

Conclusion

Information has become a primary factor in arable farming. Decision support systems for process optimisation, reporting requirements and benchmarking tools and communication in general require a widely accepted meta-model for information and knowledge in this field. We have accumulated a fairly comprehensive collection of information elements from various sources for constructing a shared conceptualisation or *ontology*. Ontologies pave the way for future generations of self-learning systems and semantic web services. We have constructed the first version of the Crop Ontology using the web ontology language OWL.

Presently, the ontology focuses on concepts such as crop, parcel, treatment, soil. In particular it covers arable farming of *potato*.

We plan to release the ontology to experts world-wide once it is sufficiently stable. Moreover, a supervised process of updating the ontology will be installed based on feedback obtained. In addition, the possibility and usability of associated web services will be investigated in order to bring the ontology to its full potential.

References

Achten VJTM, Jansen DM, Verdouw CN, Molema GJ (2005) Kennis op de akker (KodA): verkenning van kennis en mogelijkheden op het gebied van managementondersteuning op akkerbouwbedrijven. Knowledge to the Acre (KodA): exploring knowledge and potential management support at arable farms. Research Report Wageningen University, Agrotechnology and Food Science Group, 83 pp. (In Dutch)

Baarveld HR, Peeten HMG, Schipper E, Schipper JK, Delleman J (Eds) (2003). Netherlands catalogue of potato varieties. NIVAP The Hague, p 263

Been TH, Schomaker CH (2004) A geo-referenced decision support system for nematodes in potatoes. In: MacKerron DKL, Haverkort AJ (eds) Decision support systems in potato production: bringing models to practice. Wageningen Academic Publishers, Wageningen, pp 154–167

Booij R, Uenk DF (2004) Crop reflection based DSS for supplemental nitrogen dressings in potato production. In: MacKerron DKL, Haverkort AJ (eds) Decision support systems in potato production: bringing models to practice. Wageningen Academic Publishers, Wageningen, pp 46–53

Dekkers WA (2002) Kwantitatieve informatie, akkerbouw en vollegrondsgroententeelt (In Dutch, Quantitative information for arable farming and field production of vegetables). PPO-Lelystad, p 331

De Laat PJM (1980) Model for unsaturated flow above a shallow water-table. Agricultural Research Reports 895, Pudoc, Wageningen

FAO, AGROVOC, http://www.fao.org/aims/index.jsp

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

General Regulations Fruit and Vegetables Cologne (2004) Germany FoodPlus GmbH, Version 2.1-October. Cologne

Inmon WH (2005) Building the data warehouse. Wiley, New York, p 543

Jena (2006) See jena.sourceforge.net

Kempenaar C, Van den Boogaard R (2004) MLHD, a decision support system for rational use of herbicides: developments in potatoes. In: MacKerron DKL, Haverkort AJ (eds) Decision support systems in potato production: bringing models to practice. Wageningen Academic Publishers, Wageningen, pp 186–196

Kimball R, Ross M (2002) The data warehouse toolkit: the complete guide to dimensional modelling. Wiley, New York, p 436

Koenderink NJJP., Hulzebos JL, Rijgersberg H, Top JL (2005) Food informatics: sharing food knowledge for research and development. In: EFITA/WCCA 2005, Lisbon, Portugal, p 355

Kooman PL, Fahem M, Tegera P, Haverkort AJ (1996) Genotype environment interactions in potato 1. Light interception, total dry matter production and tuber yield. Eur J Agron 5:193–205

Korotkiy M, Top JL (2006), Onto-SOA: from ontology-enabled SOA to service-enabled ontologies. In ICIW'06, Guadeloupe

MacKerron DKL, Haverkort AJ (eds) (2004) Third International Potato Modelling Conference Dundee, March 2–5 2003. In: Proceedings as a Wageningen Academic Publisher book. Decision support systems in potato production: bringing models to practice p 238

Makkink GF (1957) Testing the Penman formula by means of lysimeters. Journal of the Institute of water engineers 11:277–288

Opticrop (2005) Manual CROP Gewis, optimizing timing and dosage of sprayings. Opticrop B.V. Wageningen, Netherlands, p 16

Protégé (2006) http://protege.stanford.edu/, version December 2005

Raatjes P, Hadders J, Martin D, Hinds H (2004) Plant-plus: turn-key solution for disease forecasting and irrigation management. In: MacKerron DKL, Haverkort AJ (eds) Decision support systems in potato production: bringing models to practice. Wageningen Academic Publishers, Wageningen, pp 168–185

Seamless (2006) http://www.seamless-ip.org/

VAVI (2004) Grower handbook food safety certificate for the potato processing industry 2004/2005 growing and storage season. VAVI the Hague, p 16

Verdenius F (2005) Methodological aspects of designing induction-based applications, PhD thesis, University of Amsterdam, http://dare.uva.nl/record/146542

Verdenius F, Top JL (1998) Case based modelling of dynamic systems. In: Tijskens LMM, Hertog MLAM (eds) Proceedings of the first international symposium MODEL-IT, Wageningen. Acta Horticultura 476:279–288

Vereecken H, Maes J, Feyen J, Darius P (1989) Estimating the soil moisture retention characteristics from texture, bulk density and carbon content. Soil Sci 148:389–403

World Wide Web Consortium (2006) http://www.w3c.org

Wösten JHM, Finke PA, Jansen MWJ (1995) Comparison of class and continuous pedotransfer functions to generate soil hydrolic characteristics. Geoderma 90:169–185

Wolfert J., Schoorlemmer HB, Paree PMG, Zunneberg W, van Hoven JPC (2005) Kennis op de Akker – Een programmeringsstudie Knowledge to the Acre, a programming study, LEI report 2.05.05, The Hague.