Project title:	Development of a Soil Management Information System (SMIS)
Project number:	CP107D
Project leader:	R J Rickson, Cranfield University
Report:	Annual report (24 months), November 2017
Previous report:	Annual report (12 months), December 2016
Key staff:	Rob Simmons, Tomas Kurowski, Fady Mohareb, Caroline Keay, Jason Carvalho, Joanna Niziolomski, Lynda Deeks, Jane Rickson
Location of project:	Cranfield University
Industry Representative:	Guy Thallon, FarmCare, Longside Barns, Unit 8a, Jebb Lane, Haigh, Barnsley, South Yorks, S75 4BS
Date project commenced:	01/11/15
Date project completed (or expected completion date):	31/10/18

DISCLAIMER

AUTHENTICATION

We declare that this work was done under our supervision according to the procedures described herein and that the report represents a true and accurate record of the results obtained.

[Name] R J Rickson

[Position] Professor of Soil Erosion and Conservation

[Organisation] Cranfield University

Signature .... Date ....31/10/2017........................

[Name] R W Simmons

[Position] Senior Lecturer in Sustainable Soil Management

[Organisation] Cranfield University

Signature .... Date ....31/10/17..........................

Report authorised by:

[Name] Professor A Thompson

[Position] Head of Cranfield Soil and Agrifood Institute

[Organisation] Cranfield University

Signature .... Date ....12th December, 2017.....

[Name]

[Position]

[Organisation]

Signature .... Date ....

CONTENTS

Grower Summary

Appendix 1. SMIS project Gantt chart summarising Milestones and Deliverables 41
Appendix 2. Literature review report 42
Appendix 3. Soil management causes and solutions: Soil borne diseases 55
Appendix 4. Relevant literature colour coded and categorised. 56
Appendix 5. 2015 rotational context of vining pea growers 105
Appendix 6. Report from the SMIS Stakeholder Workshop, June 26th 2017 10.00 – 14.30. Cranfield University. 108

GROWER SUMMARY

1.1. HEADLINE

The Soil Management Information System (SMIS) can add significant value to grower data by linking this to other sources of knowledge to identify the drivers behind soil management problems and their solutions. SMIS will deliver commercial benefits to UK horticulture by identifying more effective practices of soil and crop protection.

1.2. BACKGROUND

Soil management is at the heart of sustainable intensification as it has the potential to improve crop yield (both quantity and quality), whilst protecting soil and water resources. In 2013, AHDB Horticulture commissioned a gap analysis of soil management research and knowledge transfer in horticulture to inform future research programmes (CP107). Incorporating growers' views and requirements (**Table 1**), the final report identified a number of gaps in the research evidence, including the limitations of the experimental empirical base and the need for 'big data' approaches, especially given the unprecedented amount of data being generated by growers through on-farm data management software such as 'Gatekeeper'. Indeed, many growers already collect data on many aspects of crop agronomy, field operations and soil health as part of their routine farm management. While some of it is used for business planning or to support assurance and certification schemes, there is under-utilised potential that could be used to optimise benefits on farm. Some of this data has the potential to enhance the productivity and competitiveness of growers' businesses, including data that could support sustainable soil management or drive innovation in cropping systems. However, these potential benefits can't be realised from data from one business on its own or even a few businesses working together.

Also, it was recognised that sources of information and data related to soil management in horticulture are currently unstructured, uncentralised and difficult to find and/or access. A real opportunity exists to optimise the integration of diverse sources of information pertaining to soil management issues in horticulture, and their effective solutions. As a result, it was recommended that future research should develop a soil management information system (SMIS) that will hold, manipulate and manage such data in a way that can be interrogated to provide advice and guidance on the benefits of soil management practices, with regard to crop productivity and environmental protection.

The aim of this 3 year project (2015-2018) is to apply the principles of 'Big Data' to provide best practice guidelines for sustainable soil management in horticulture.

Table 1. Soil management issues identified in CP107 (Rickson and Deeks, 2013)

	HDC Panel (2013)							
Soil management issue	Field Vegetables	Bulbs & Outdoor flowers	Soft fruit	Protected edibles	Protected ornamentals	Tree fruit	Mushrooms	Hardy nursery stock
Increased productivity	✓	✓	✓	✓	✓	✓	✓	✓
Control of pests, diseases, weeds and volunteers	✓	✓	✓	✓	✓	✓		
Use of automation /precision agronomy /smart mechanisation	✓	✓		✓	✓			
Improved monitoring techniques	✓			✓			✓	
Surface/subsurface water management			✓	✓	✓	✓		✓
Control of environmental impacts	✓	✓		✓	✓		✓	✓
Use of composts, mulches, green wastes, green manures		✓			✓	✓		
Use of alternative growing media	✓		✓		✓			✓

1.3. SUMMARY

The purpose of the Soil Management Information System (SMIS) is to gather and collate diverse sources of information and data on soil management issues and their mitigation, and harmonise it. The integration of wide-ranging data forms and sources makes SMIS unique: such a holistic approach has never been attempted before. The overall database so created can be interrogated with data mining techniques (i.e. the SMIS Analytical Toolkit) to establish patterns in the data. This ‘rule base’ establishes the relationships between the intrinsic factors that affect soil condition (e.g. site conditions, soil type, crop, weather, timing of operations, machinery) and best management practice for soil management operations, together with anticipated outcomes of given actions (for all SMIS data sources). The aim is to identify the drivers behind key soil management issues for horticultural growers such as soil borne diseases and soil degradation.

SMIS will give growers, agronomists and land managers access to guidance and information on evidence-based, optimal soil management practices in horticulture. Critically, as the size and number of records in the database increases, SMIS can learn and reincorporate new data as it is introduced in the form of updated probabilities and likely outcomes.

Year 1 of the project focussed on the collation of data, information and knowledge on soil management issues and their solutions from a wide range of sources, including farmer/grower data (e.g. as held in Gatekeeper farm management software); other relevant datasets (e.g. Met Office data; LandIS Soil Series data); research outcomes from on-going AHDB-

Horticulture funded projects (i.e. CP107b and CP107c); expert knowledge from researchers and growers; and an extensive review of the literature (both scientific papers, research reports and 'grey' articles).

However, the certainty and confidence in the rules / relationships developed within SMIS depends on the extent and quality of the database. The project team continues to engage with growers to access more anonymised grower data that will generate and validate the relationships generated by the SMIS. Better engagement with growers over the year has led to a number of datasets to be uploaded to the database.

Where data has been supplied, unique merging of grower data with the LandIS database on soils and meteorology has been used to identify risks of compaction and structural degradation (e.g. slaking, surface capping), windows of opportunity for non-soil damaging field operations (workability days). Such insights would have a direct commercial benefit to UK horticulture (see Financial Benefits below). Data-rich AHDB projects such as the FV5 cavity spot projects (1990 – 2007) have been included in SMIS and linked with LandIS data to explore new relationships between treatments that were not visible in the original experimental trials.

However, accessing anonymised grower data is still a challenge. There are a number of reasons for this: heavy workloads throughout the year; limited time 'in the office' to access the data; concerns about the anonymity of the data; concerns about how clean and extensive the data might be – and whether it is any use to the project, especially if data is not recorded. Although the protocol for uploading data is self-explanatory (Appendix 11), many growers prefer to meet in person to go through their data records. This can be time consuming and is not sustainable in the long term (i.e. post project).

Incorporating information from the literature has expanded considerably in Year 2 (2017), with screening of 581 individual items of literature with broad coverage of horticultural sectors (although only 84 were within the scope of SMIS' aims). The knowledge from the literature has been categorised to reflect certainty and confidence in these sources. This categorisation will inform the strength of relationships in the SMIS rules base (e.g. soil management practices and effects on yields, soil health etc.). Linkage with other AHDB GREAT soils projects has the potential to add significant value to the SMIS dataset, if the soil metrics measured can be critically evaluated against field operations that have been undertaken on the fields sampled across their full rotational context.

We have received positive industry engagement, including their understanding of potential benefits and contributions to the development of the project. The outputs of the project will be of direct benefit to levy payers because:

- For the first time, currently unstructured, uncentralised and difficult to find and/or access information on soil management advice and guidance in horticulture will be available as one centralised resource;
- Storing, cleaning, transforming and analysing such large and complex collections of different kinds of data is beyond the normal computing capacity of most individual businesses. New and emerging methods of data management and processing known as ‘agri-informatics’ will allow meaningful interpretation of such datasets to unearth valuable insights that would otherwise have remained hidden. These insights can then be presented to growers and their agronomists to inform future soil management decisions, such as improving soil health (with associated improvements in crop production) and avoiding soil degradation (and associated remedial costs).
- For individual growers, patterns in their own farm data are often obscured by the variations in soil management practices and their effects from season to season, year to year and from field to field. This ‘noise’ starts to fade as the pooled dataset gets bigger. Patterns that aren’t visible in an individual data set are more likely to be revealed and can be used, for example, as the basis for best practice guidance on soil management.
- Obtaining coherent datasets that encompass full rotations is critical to identify positive or negative effects of particular operations/practices on yield, yield quality and to promote sustainable soil management.
- Additional benefit can also be drawn from the grower data by looking at the full rotational context and the operations associated with it. For example, we may be able to identify the optimal rotations and farming practices to mitigate soil borne pests and diseases and so help the industry to reduce reliance on chemical crop protection products.

1.4. FINANCIAL BENEFITS

The soil management advice and guidance given by the SMIS will bring financial benefits for levy payers in two ways. First, by identifying the causes of soil degradation (and practices used effectively to control them), SMIS will help reduce costs incurred by growers from the impacts of soil degradation. Conservative estimates of the impacts of soil degradation on agricultural production in England and Wales alone are estimated at £212-270 million per annum (Graves et al., 2011; 2015). Soil degradation has financial consequences for individual growers both on-field and off-farm. Poor soil quality (e.g. compacted soil) leads to gaps in production continuity and critically to pinch points in product delivery. Such continuity gaps can exert significant financial impact on growers and increase the reliance on imports to meet customer requirements and to maintain national food security. Costs to individual farmers/growers may include reseeding operations, subsoiling to alleviate compaction, relevelling land subject to erosion, fines incurred due to breaches of the Water Framework Directive (eroded soil in watercourses) or from the Highways Agency (mud on roads), additions of organic amendments, and poor yields. Loss of customer confidence due to the difficulty of delivering to time and specification (quantity and quality) can also have a significant longer-term impact on farm income. The SMIS database can be used to identify how the timing of tillage, planting and harvesting operations are linked to soil degradation which in turn affects yield and yield quality. This information will inform and justify future soil management decisions that avoid soil degradation.

Second, SMIS will provide more effective advice and guidance on soil management practices that lead to improved soil health and system resilience. As well as increased outputs (yield quantity and quality; **Table 2**), well managed soils have lower input requirements (nutrients, water, agrochemicals), giving better financial margins in the short term, and better soil quality / health in the long term.

Table 2. Increase in yields of crops grown in horticultural rotations due to improved soil and water management (2015 prices)

Crop	Yield increase associated with better soil health	Financial benefits to individual growers
Wheat	up to and over 10%	10% increase in yield would result in 1.2 t/ha increase @ £130/t
Potatoes	5%	based on 15,000t produced = 750t extra – contract price £165 /t = £123,750 income
Maize	5%	Improved yield means less land required. If 40 ha of land under maize @ growing cost per ha of £1550k = saving of £65,000. The 40 ha could be put to wheat = 528 tonnes = £68k income.
Lettuce	1.5%	Improved yields mean 1.5 million fewer heads per yr needed = 15 ha less land @ growing cost per ha of £8k = saving of £120,000
Onions	2.5%	based on 5000t produced = 125t extra yield – contract price £190 per ton = £23,750 income

1.5. ACTION POINTS

To achieve the benefits of SMIS (improved soil management advice and guidance, (e.g. enhancing crop production and environmental protection)), growers are asked to:

- Provide data and information to the SMIS data repository that contain soil metrics and field operations undertaken on fields sampled across their full rotational context. As highlighted by Martin Evans and John Chinn at the SMIS stakeholder workshop (June 2017; Appendix

6), it is especially important that we bring in data for a larger number of field vegetable growers representing multiple grower associations.

- Provide field records other than Gatekeeper (such as Muddy Boots). It is recognised that other data capture software and records exist and maintain a significant share of users among growers, but they are not represented among the datasets used during SMIS development. Initial SMIS development is based on growers' data collected in Gatekeeper format. As SMIS is intended to be extendible and capable of being adapted to handle disparate sources of grower data, we will consider other data formats, but within the constraint that the SMIS backend architecture has now been finalised.
- Identify the queries to be run in SMIS. Although the previous gap analysis and grower survey (Rickson and Deeks, 2013) identified key soil management issues in the horticultural sector, it is recognised that these may have evolved in the intervening 5 years. Growers should identify the key challenges facing their businesses in terms of soil management. One example is the opportunity to optimise the use of LandIS derived outputs and participating grower data (Figure 9) to augment the data obtained from FV373 and use agri-informatics data mining approaches to explore soil management drivers of cavity spot incidence and severity. Agri-informatics data mining approaches will then explore the soil management issues identified.

SCIENCE SECTION

1. INTRODUCTION

Currently, the horticultural sector lacks access to coherent information to support decisions on sustainable soil management. This is despite the unprecedented growth in agricultural data. There is a need to identify analytical approaches that determine differences between sustainable and unsustainable sites / practices, and best practice guidelines for sustainable soil management in horticulture. This is not possible currently, due to uncertainty surrounding the evidence base. Although some knowledge related to soil management in horticulture does exist, it is dispersed throughout the sector. It is often limited by the scale and duration of studies (e.g. range of soil types, crops and/or management practices). It often concentrates on individual crops (rather than considering the longer term, rotational context) and there is little reporting on the cost effectiveness or practicality of the measures used.

The aims of the SMIS project are to apply the principles of ‘big data’ to the diverse sources of soil management data, knowledge and information. The SMIS will hold, represent, manipulate and manage available sources of data, knowledge and information, with specific focus on the effects of soil management practices on horticultural crop productivity and environmental protection. SMIS operates over a seasonal and (more innovatively and uniquely) cross-rotational timeframe, allowing legacy effects of previous soil management decisions to be captured. It also links grower (anonymised) data to the LandIS soil and environmental datasets. By improving the evidence base, SMIS will inform on-farm decisions on horticultural soil management.

Novel informatics techniques are used to extract patterns of ‘cause and effect’ regarding soil management practices (and their outcomes) in different scenarios (e.g. soil type, crop, rotation, location, etc.). This is the SMIS ‘rule base’, which can then be interrogated with specific queries related to soil management issues. These might include how to: increase crop yield and quality; minimise environmental impacts, including soil degradation; reduce the cost of mitigation / remedial measures; allow crops to be drilled and harvested at optimum times; maintain good soil condition; increase marketable yield; extend cropping seasons without environmental damage; and/or improve system resilience to climate extremes.

The outcome will be a set of robust, empirically-based, best-practice soil management guidelines (and the likely consequences of applying them). An interactive platform has been developed, giving AHDB-Horticulture, and its growers, agronomists and land managers access to guidance on contextual, effective soil management practices.

1.1. Actions points from Year 1 of the SMIS project

The following action points were identified in the SMIS Annual Report (2016). Year 2 (2017) activities to address these are shown in Table 3 and detailed in the relevant sections below.

Table 3. Action points from Year 1 Annual Report and corresponding activities in Year 2.

Action Points from Year 1 Annual Report (2016)	Activities in Year 2 (2017)
To achieve the benefits of improved soil management advice and guidance, (e.g. enhancing crop production and environmental protection), it is necessary to provide data and information to the SMIS data repository. We will continue to source additional inputs to SMIS by working closely with our industrial partners, project collaborators (e.g. PGRO), other researchers (especially those undertaking CP107 projects) and developing additional contacts in the horticultural sector.	Data collection (See Section 2.1 and 3.1)
Inputs will be combined with other datasets such as Met Office data and LandIS Soil Series data to present specific soil management issues faced by growers, such as soil compaction.	Integrating grower data with LandIS and Met Office data (Section 2.1.3. and 3.1.7)
Issues of data integrity, reliability and accuracy will be addressed as new sources are incorporated into SMIS. This will include how to manage missing data – is it possible to use proxy data instead? For example, soil bulk density measurements are not always available (as an indicator of soil compaction). However, knowing the soil type, weather conditions and machinery used, the risk of soil compaction can be estimated.	Database technical documentation (Section 3.2.1)
The large datasets within SMIS will require the use of complex data management techniques and advanced computational skills. We will continue to develop analytical methods and statistical modelling, drawing across the body of data assembled, allowing comparative assessment and benchmarking against available grower and case-study data. Statistical interpretation of grower datasets within SMIS will provide a more scientific basis for guidance on a wide range of soil management issues.	Building the SMIS backend (Section 3.2) and User Interface and Analytics Toolkit (Section 3.3)
A rules-base for functional relationships between data members will be established based on expert opinion, established AHDB guidance documents and weight of evidence in the literature. This will form the basis of a suite of expert knowledge and hypothesis driven statistical analyses.	Building SMIS backend (Section 3.2) and User Interface and Analytics Toolkit (Section 3.3)
As the database within SMIS develops, it is envisaged that data mining techniques will provide useful insights to address AHDB Horticulture Panels' 2015-2018 priorities.	Building the SMIS Backend (Section 3.2) and SMIS User Interface and

	Analytics Toolkit (Section 3.3)
We will continue to promote SMIS in project knowledge exchange activities.	Knowledge and technology transfer (Section 6)

2. MATERIALS AND METHODS

The structure of this section follows the Project's Milestones and Deliverables (Appendix 1).

2.1. Data collection and collation

SMIS is designed to hold, represent, manipulate and manage diverse and currently disparate sources of data, knowledge and information on soil management issues, practices and outcomes. Various sources of information now populate the SMIS database, which can be interrogated with user queries using the analytics toolkit, which is currently under development.

2.1.1. Data from the literature and research reports

The literature review report (submitted on 28/04/17; Appendix 2) describes the methodology for data collection from the literature and research reports. Sources of information covered in the review included:

- Academic papers published in scientific, peer reviewed journals;
- Conference proceedings / papers;
- Research reports;
- Grey literature (e.g. articles on websites and in trade magazines)

Collectively, these sources are termed 'literature' in the following sections. Each source (or 'item') was classified as being 'quantitative', 'qualitative' and 'anecdotal' (see Appendix 4). This classification was used to evaluate and quantify the confidence in outputs / findings from each item (i.e. the 'weight of evidence' within the SMIS database). It is envisaged that a weighting system needs developing to reflect the confidence behind all the different data sources within SMIS (i.e. grower data [Gatekeeper software], literature, research projects, expert opinion: See Annual Report, November 2016 for further detail of information / data / knowledge sources to be used in SMIS). The Cranfield team will consider how an evidence weighting system can be incorporated into the SMIS system architecture (e.g. through visualisation).

The review of the literature had two main phases (Figure 1). The first phase required the collation of literature relevant to horticultural soil management. This brought together literature from two sources:

- the literature identified in the gap-analysis review conducted by Rickson and Deeks (2013).

- a new literature search (detailed below) that updated Rickson and Deeks' (2013) literature with relevant published papers between 2013 and 2016.

The second phase was then to review the collated literature from both sources in order to identify, categorise and catalogue case studies that had potential in providing knowledge that can be incorporated into the SMIS database (Figure 1). Results are presented in 3.1.1 below.

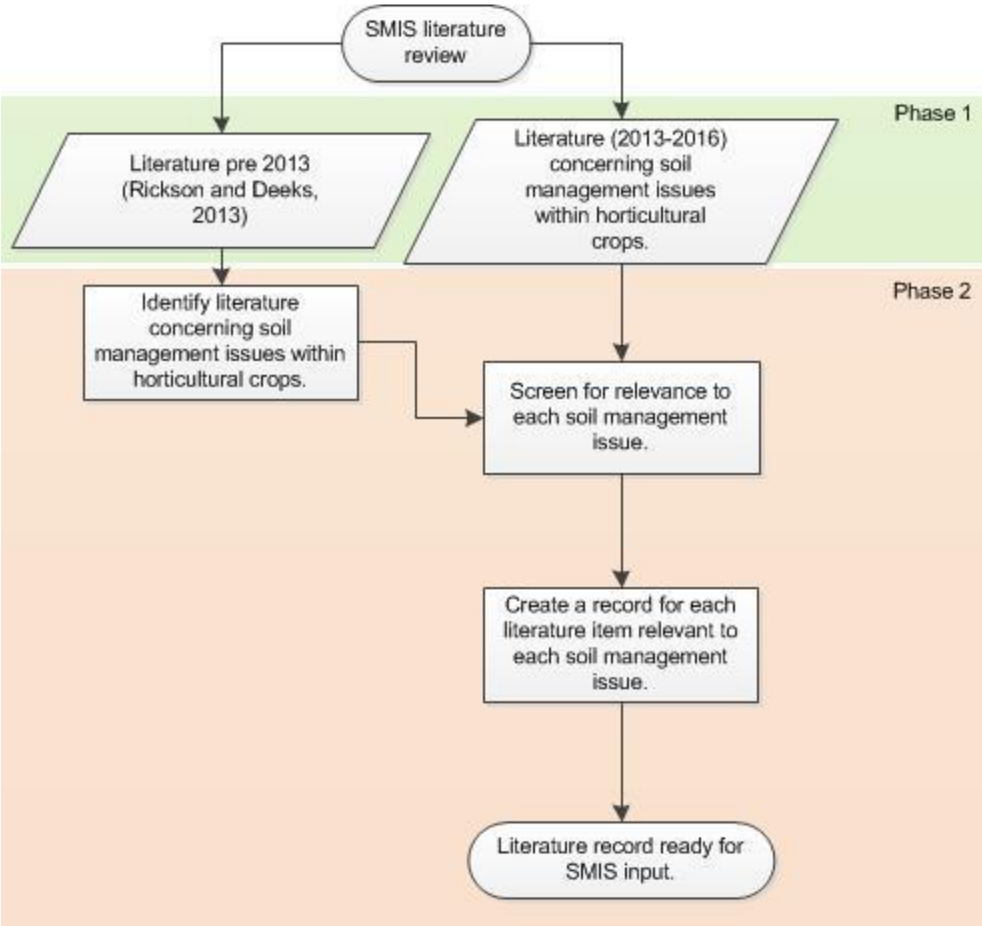


Figure 1. Literature review process

2.1.2. Grower data

Farmers and growers are collecting significant amounts of data from their field and farm operations. Much of this is pertinent to soil management challenges such as soil compaction, which can directly impact on crop yield, financial returns and environmental quality. Data is also available on the practices that have been used to prevent or remediate these issues. The methodology for extracting grower data to populate the SMIS database is described in Appendix 11**.** The grower datasets collected to date are reported in section 3.1.4 below.

2.1.3. Integrating grower data with LandIS, RPA field boundaries and Met Office data

2.1.3.1. Background: How LandIS data is used to address soil management problems

a) Workability and trafficability

Workability (Thomasson 1982) of the soil depends on interactions between climate and soil physical properties. For example, good working conditions on clayey soils are commonly restricted to brief periods when the soil is neither too wet nor too dry for a good tilth to be obtained. Similarly, trafficability of land by machinery and stock is restricted by soil wetness. The influence of topsoil properties, permeability and wetness is integrated to give assessments of workability and trafficability which are applied to the main arable soils. For example, assessment aa is used for very coarse well drained soils and assessment f is used for extremely wet, heavy or peaty soils. The restriction on workability and trafficability increases from aa to f. The wetness class of the soil is used on the assumption that, where appropriate, feasible drainage improvements have been undertaken.

On much land, the field capacity period sets broad limits to good ground conditions for tillage and trafficability with conventional machinery. The median return dates of the field capacity period are used to estimate the limits of the landwork periods in a normal year. Related quartile dates are used similarly for wet years which occur with a frequency of one in four. From these dates, and the soil assessments for workability and their weightings, potential machinery work days (Smith 1977) are calculated to provide a measure of the number of days when the land can be worked with acceptable risk of damage to soil structure during the main autumn and spring activities of harvesting, tillage and drilling.

b) Predisposition to structural damage

Repeated arable cropping can progressively degrade soil structure especially if good soil management is not practised. Even one badly mis-timed operation (at crop establishment or during harvesting) can seriously degrade soil structure. Soil structural degradation can take the form of:

- soil surface capping,
- soil compaction, and
- weakening aggregation and hence changing the shape and size of aggregates.

These degradation features can result from over working the soil (especially with power harrows), reduction in organic matter content or from undertaking mechanical operations (establishing, fertilising and harvesting crops) when the ground is wet.

Soil properties that influence structural changes are Wetness Class, particle-size class (texture), organic carbon content and whether the soil is calcareous or non-calcareous.

Current best estimates have been provided for three elements of structural degradation and remediation. In each case for the following schemes a 4-tier ranking of structural stability for arable soils is given:

• Structure susceptible to Topsoil Slaking, (StoTS) classes 1 to 4: very unstable, unstable, moderately stable and stable;

• Structure susceptible to Compaction (StoC) classes 1 to 4: very susceptible, moderately susceptible, slightly susceptible and very slightly susceptible;

• Potential of soil for Natural Structural Regeneration (NR) classes 1 to 4: little potential, slight potential; moderate potential and large potential.

The results of this activity are presented in section 3.1.7 below.

2.2. SMIS Back End

The set of software which together form a group of key deliverables of the Soil Management Information System (SMIS) project underwent a significant redesign during the course of the year. The changes included a reformulation of the software project structure and an explicit articulation of specific assumptions about how the completed system will operate, what technologies and frameworks it will depend upon, what its requirements will be, what types of Use Cases it will support, and what sort of results it will produce.

In particular, an overall modular structure establishing the individual software components and their relationships was described and agreed upon. A simplified diagram showing the information flow in the system can be seen in Figure 2. General diagram of information flow in the SMIS system.The fundamental elements (described in more detail in individual submitted reports) comprising the SMIS software deliverables are:

- Parsing tools for collected data
- Non-relational SMIS database
- REST API back-end for the database
- Analytics tools supporting collected data analysis
- SMIS Analytics Toolkit web application providing a front-end for the other components

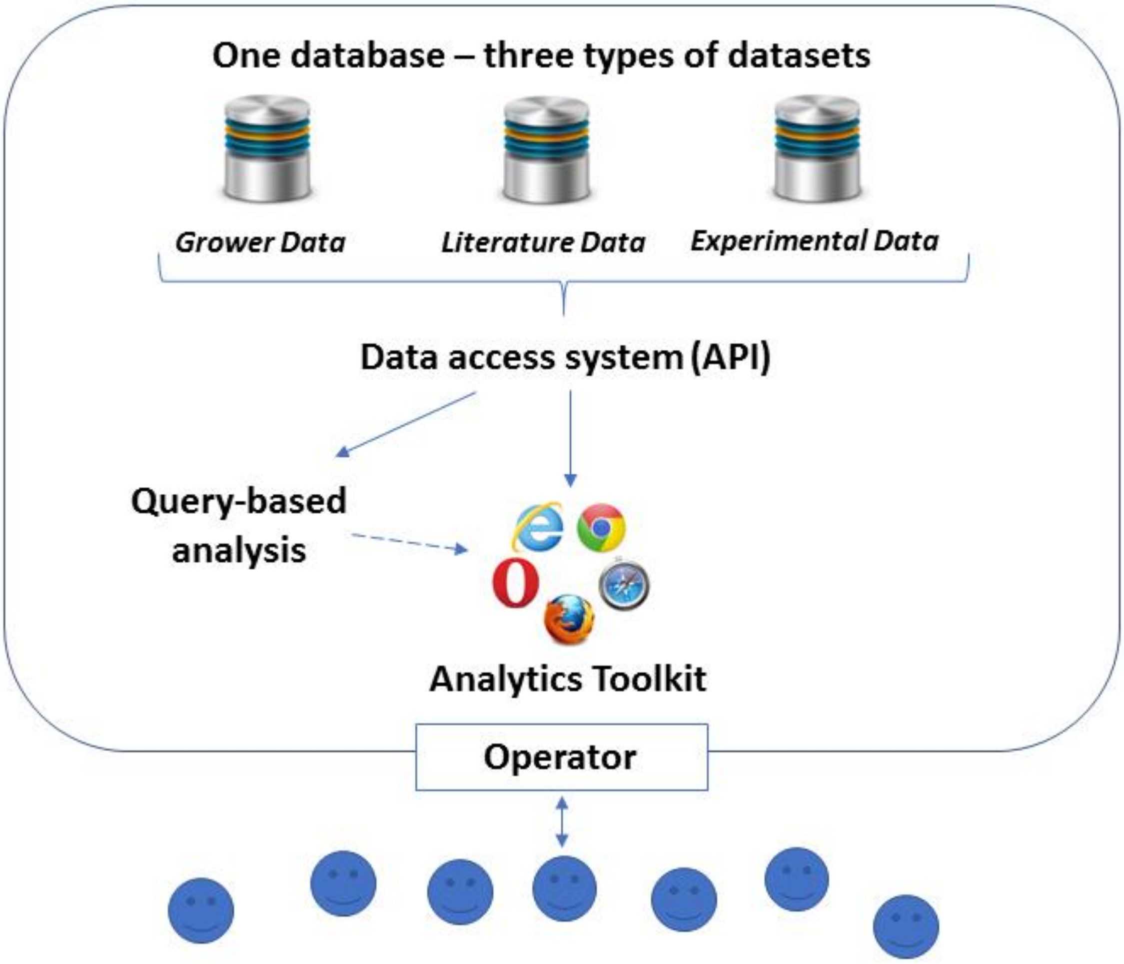


Figure 2. General diagram of information flow in the SMIS system.

This work was articulated in two reports formulated in August (SMIS Database Architecture Technical Documentation – Appendix 9) and September 2017 (SMIS Use Case Documentation – Appendix 10). Both documents are intended to serve the dual purpose of providing an overview of the design and an accurate description of the technical details of the system to the end user, as well as serving as a specification and implementation guide for the developer.

3. RESULTS

3.1. Data Collection and Collation

3.1.1. Data from the literature and research reports

Since the 2016 annual report, the initial review of literature has been completed (reported separately to AHDB Horticulture 1st May 2017). Consequently, a more complete analysis of the literature identified is available.

In total, 581 items of literature were reviewed. These included:

- Academic papers published in scientific, peer reviewed journals;
- Conference proceedings / papers;
- Research reports (e.g. AHDB, Defra);
- Grey literature (e.g. articles on websites and in trade magazines)

Of these, 84 items hold knowledge that are directly relevant to SMIS aims. Due to scale and relevance, laboratory studies and results from non-temperate climates were excluded from the final list.

Within the 84 items, 172 individual knowledge items for horticultural soil management were found. This larger number results from items that address multiple soil management issues and/or practices. For example, Bonanomi et al. (2008) addresses the control of weeds in protected edibles using a number of soil management practices: the use of soil solarisation, plastic mulches and soil amendments.

Each item of literature was classified by knowledge type; quantitative (based on empirical evidence from field work: laboratory studies were excluded due to the limitations of extrapolating practical, applied results from small spatial scales); qualitative (based on observations during a field-based experiment); and anecdotal (unreferenced statements). This classification was used to evaluate and quantify the confidence in outputs / findings from each item (i.e. the 'weight of evidence' within the SMIS database). It is envisaged that a weighting system needs developing to reflect the confidence behind all the different data sources within SMIS (i.e. grower data [e.g. Gatekeeper], literature, research projects, expert opinion: See Annual Report, November 2016 for further detail of information / data / knowledge sources to be used in SMIS). It was assumed that items with quantified data would provide more confidence to end users than qualitative or anecdotal information. Therefore, for each soil management issue, the specific details of available quantitative knowledge within each item of literature were extracted into a common descriptive form (termed meta-criteria in the Annual

Report 2016). Quantitative references to soil management challenges were the most frequent, making up 61 % of the identified knowledge sources (Figure 3).

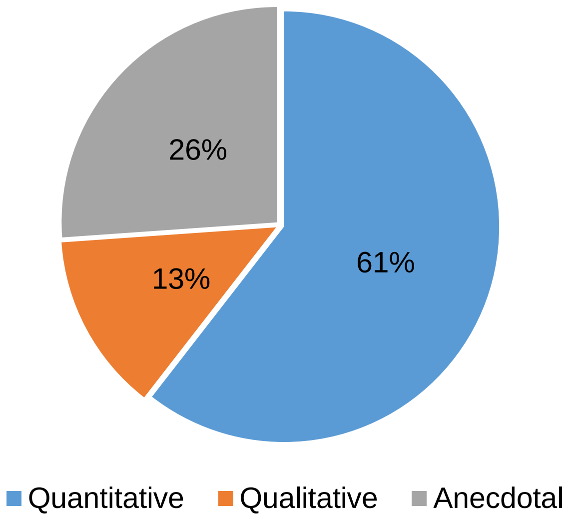


Figure 3.Classification of knowledge sources within the relevant literature.

The greatest number of knowledge items focus on research undertaken in the UK, with a good global distribution of other literature (Figure 4). Field vegetable, tree fruit, cross sector and protected edibles are particularly well represented (Figure 5). The greatest amount of literature was found for soil-borne disease, followed by weeds and nutrient supply (Figure 6). The literature covers 20 broad soil management solutions with a focus towards generic 'management practice' (organic versus conventional) and rotation based solutions (Figure 7). All other solutions are evenly distributed across the identified literature.

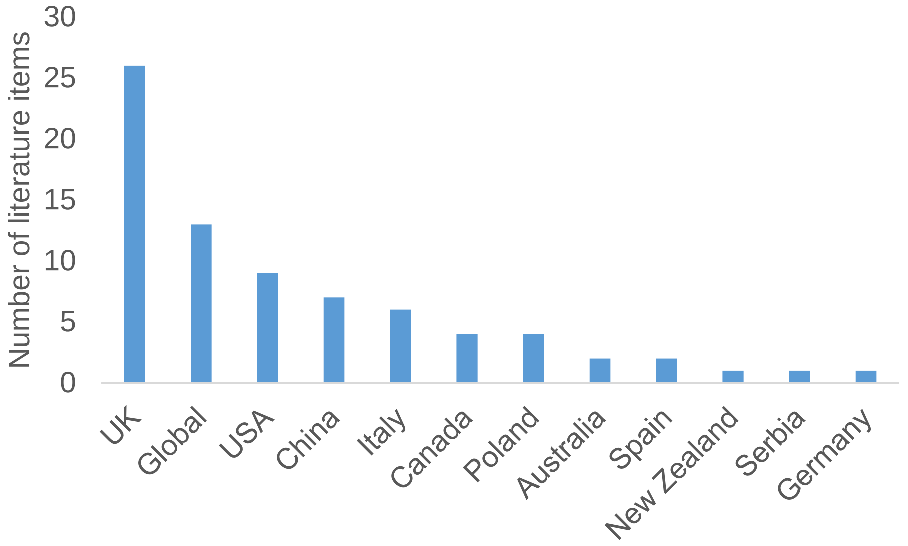


Figure 4. Distribution of relevant literature by country.

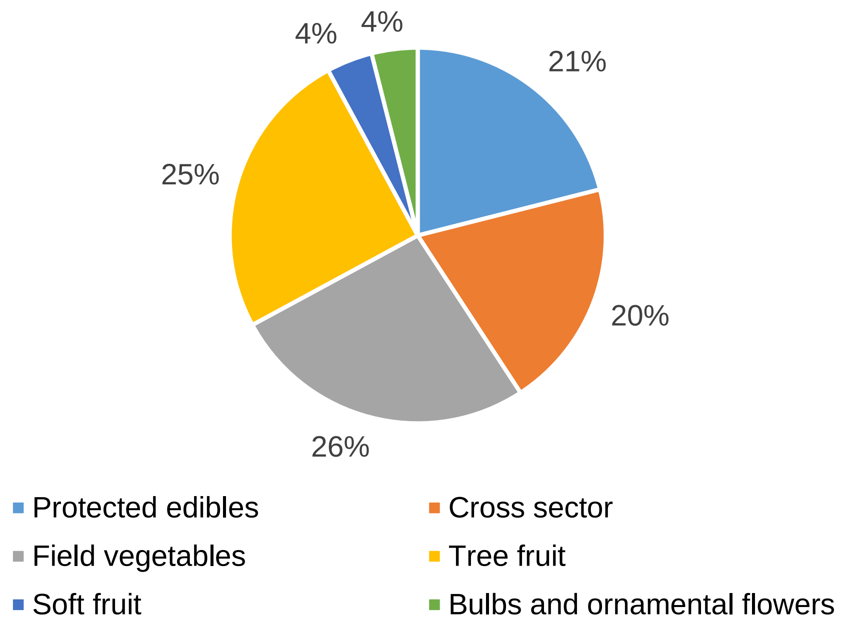


Figure 5. The distribution of relevant literature by AHDB Horticulture sector.

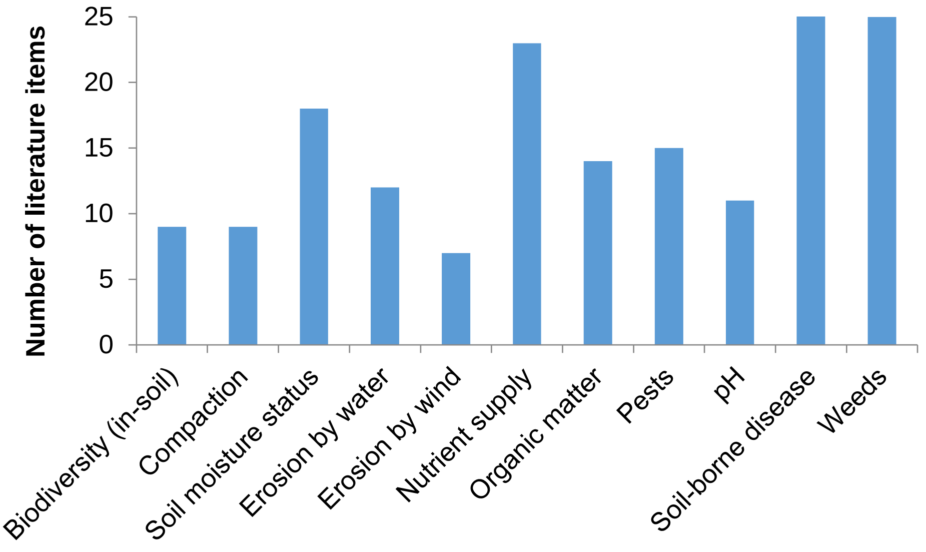


Figure 6. Distribution of knowledge items across each soil management challenge.

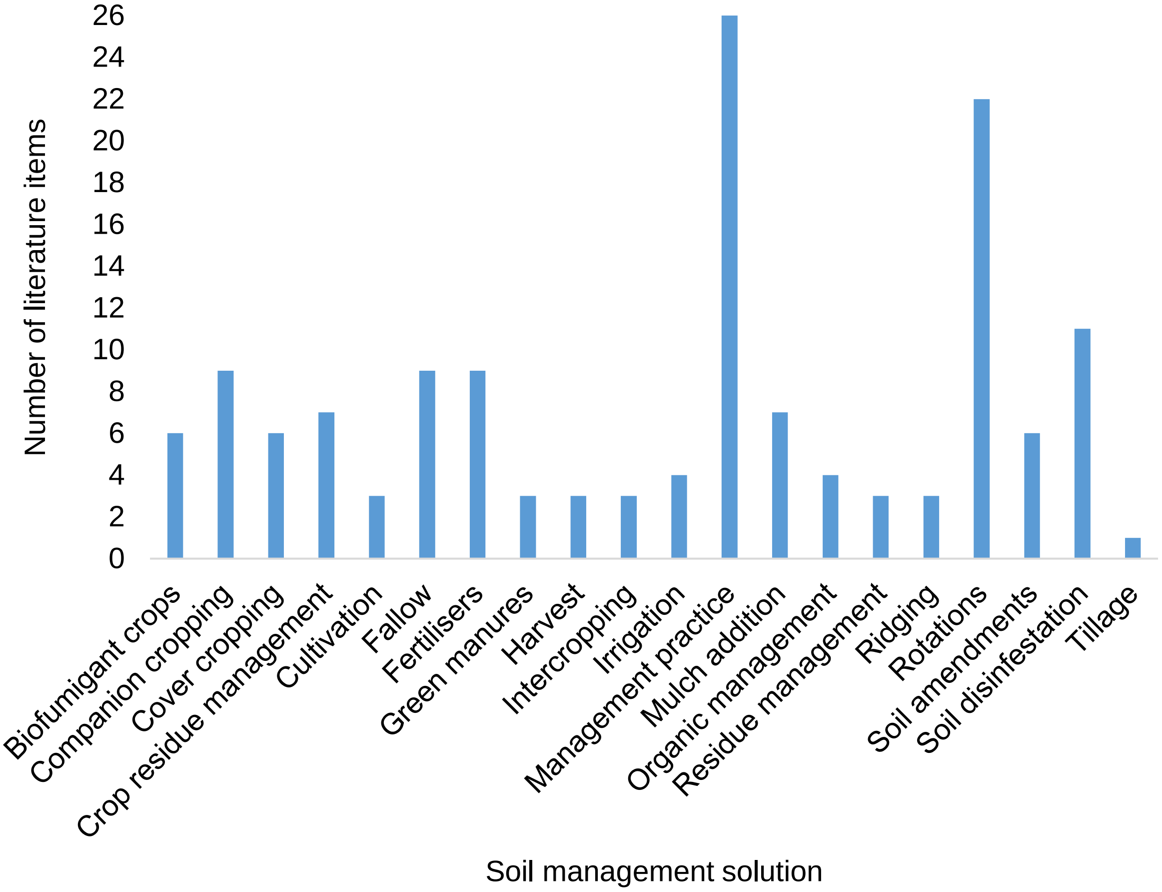


Figure 7. Distribution of knowledge items relative to soil management solutions.

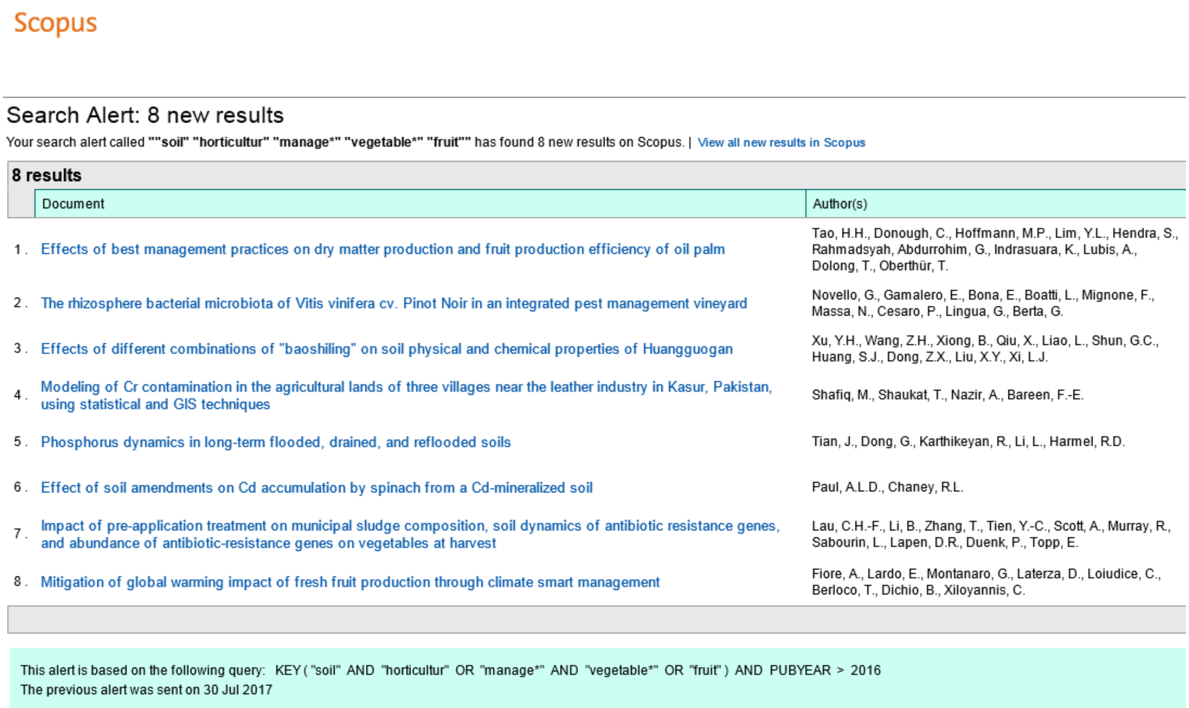
The literature identified in this review forms a sound and integrated basis of horticultural soil management knowledge. Benefits to the SMIS end users include:

- a unique, novel and up-to-date synthesis of the extensive and diverse research outputs related to horticultural soil management issues and solutions
- easy access to knowledge that has previously been inaccessible (e.g. locked up in subscription-based academic papers)
- expansion of the SMIS database in terms of quantitative, qualitative and anecdotal knowledge on both horticultural soil management issues and management practices used to both prevent and remediate these issues.

Having extracted the relevant information from the literature review in a systematic way, the next step is to work with Cranfield colleagues on the informatics team to understand how the knowledge identified in the literature review (in quantitative, qualitative and / or anecdotal form) can be accommodated in the SMIS database, alongside the grower data (e.g. Gatekeeper records) and expert knowledge/ opinions (e.g. outputs from the Fuzzy Cognitive Mapping exercise). The vision is for the resulting SMIS database to be interrogated to identify best soil management practices to:

- alleviate and prevent soil degradation in horticultural crop production
- reduce input costs associated with remediation of soil management problems in horticulture
- remove the soil-based constraints that currently limit optimal crop production in horticulture; and
- enhance soil quality and soil health to bring agronomic and financial benefits to the grower.

Since the review of literature relevant to SMIS has been completed (reported to AHDB Hort. on the 1st May 2017) the literature review has been regularly revisited for any available updates with new literature. A search alert has been set up with Scopus® (http://www.scopus.com/home.url) to capture new peer-reviewed literature (including scientific journals, books and conference proceedings) relevant to horticultural soil management (Figure 8). New research projects are also investigated periodically for relevance to SMIS. This includes reviewing AHDB Horticulture, Defra and BBSRC research project webpages. To date, no additional, new relevant research has been identified.



Scopus

Search Alert: 8 new results

Your search alert called ""soil" "horticultur" "manage*" "vegetable*" "fruit"" has found 8 new results on Scopus. | View all new results in Scopus

8 results

	Document	Author(s)
1.	Effects of best management practices on dry matter production and fruit production efficiency of oil palm	Tao, H.H., Donough, C., Hoffmann, M.P., Lim, Y.L., Hendra, S., Rahmadsyah, Abdurrohim, G., Indrasuara, K., Lubis, A., Dolong, T., Oberthür, T.
2.	The rhizosphere bacterial microbiota of Vitis vinifera cv. Pinot Noir in an integrated pest management vineyard	Novello, G., Gamalero, E., Bona, E., Boatti, L., Mignone, F., Massa, N., Cesaro, P., Lingua, G., Berta, G.
3.	Effects of different combinations of "baoshiling" on soil physical and chemical properties of Huangguogan	Xu, Y.H., Wang, Z.H., Xiong, B., Qiu, X., Liao, L., Shun, G.C., Huang, S.J., Dong, Z.X., Liu, X.Y., Xi, L.J.
4.	Modeling of Cr contamination in the agricultural lands of three villages near the leather industry in Kasur, Pakistan, using statistical and GIS techniques	Shafiq, M., Shaukat, T., Nazir, A., Bareen, F.-E.
5.	Phosphorus dynamics in long-term flooded, drained, and reflooded soils	Tian, J., Dong, G., Karthikeyan, R., Li, L., Harmel, R.D.
6.	Effect of soil amendments on Cd accumulation by spinach from a Cd-mineralized soil	Paul, A.L.D., Chaney, R.L.
7.	Impact of pre-application treatment on municipal sludge composition, soil dynamics of antibiotic resistance genes, and abundance of antibiotic-resistance genes on vegetables at harvest	Lau, C.H.-F., Li, B., Zhang, T., Tien, Y.-C., Scott, A., Murray, R., Sabourin, L., Lapen, D.R., Duenk, P., Topp, E.
8.	Mitigation of global warming impact of fresh fruit production through climate smart management	Fiore, A., Lardo, E., Montanaro, G., Laterza, D., Loiudice, C., Berloco, T., Dichio, B., Xiloyannis, C.

This alert is based on the following query: KEY ("soil" AND "horticultur" OR "manage*" AND "vegetable*" OR "fruit") AND PUBYEAR > 2016
The previous alert was sent on 30 Jul 2017

Figure 8. Scopus search alert output for identifying new relevant peer-previewed literature. Output dated 30/09/17.

3.1.2. Data from other CP107 projects

3.1.2.1. CP107b – GREAT soils

CP107b contains some data that could be deemed suitable for inclusion into SMIS. This data includes results from six field demonstration trials set under work package two lasting two growing seasons. These seek to investigate different soil health assessment techniques for the identification of increased soil organic matter and soil health. In doing so these trials are comparing different soil management techniques, with a variety of sites and crops selected (Table 4. An overview of the CP107b experimental trials.) making them directly relevant to SMIS.

Table 4. An overview of the CP107b experimental trials.

Site	System	Crop	Location	Treatment
1	Conventional field veg	Carrot	Scotland	Crop residue and compost application
2	Conventional protected cropping in soil	Spinach and rocket, field lettuce	West Midlands	Protected: Short-term green manures, phacelia and buckwheat

				(protected). Overwintering rye and vetch mix (field lettuce)
3	Conventional field veg	Salad	East Midlands	Short term green manure strips on beds.
4	Organic field veg	Carrot	East Midlands	Flowering green manure strips on beds.
5	Conventional top fruit	Apple	South-East	Green manure mixes x2
6	Organic field veg	Post potatoes pre brassicas	South-East	Early and late sown green manure mixtures.

It is envisaged that due to the small amount of data that will be generated from these trials that they would be incorporated as a knowledge item i.e. utilised as part of the soil management relationship weighting system and present as a detailed record that can be called upon and viewed.

Another data source from CP107b is the field labs that they are launching in Year 3. These are year-long grower led practical farm trials of which there will be 5 in total, but only the first three have been detailed (Table 5).

Table 5. Overview of the 3 out of 5 field labs forming Year 3 of the CP107b project.

Field lab no.	Title	Crops	Treatment specifics
1	Improving soil health and organic matter using cover crops in a shared rotation	Lettuce, potatoes and sugar beet.	Trial 1: Cover crops: Oats, Italian rye grass and vetch mix and a control. Trial 2: Chicken litter, compost, a mix of mustard green manure designed for PCN control and a control. Trial 3: PCN mustard mix and control.
2	Amendments for soil health in top fruit		Woodchip, woodchip as mulch, ramila woodchip compost, enriched biochar, green waste compost, digestate, mycorrhizae
3	The impact of whole digestate on soil health in field-grown vegetable crops on the Moray coast		Whole digestate from farm produced energy crops.

3.1.2.2. CP107c – Precision farming project

See section 3.1.4.3 below.

3.1.3. Other AHDB Soil Management Research Projects

AHDB Horticulture research projects have been included in the literature review and in subsequent updates. Separate to this, data-rich projects such as the FV5 cavity spot projects (1990 – 2007) have been enhanced with new soils data. These projects were found to be lacking in basic soils data, and with no significant differences observed in tested variables. With the use of geographic data collected in the reports, LandIS data such as that used within the SMIS project has been applied to consider new relationships between treatments.

3.1.4. Grower data

Wider engagement with industry has taken place in 2017, extending contacts into brassica and carrots growers (identified in discussions) and to other horticultural groups. We have received positive industry engagement, including their understanding of potential benefits and contributions to the development of the project.

3.1.4.1. Vining pea datasets

The 2016 Annual Report indicated that whole farm Gatekeeper datasets had been received from four HMC Peas growers namely Worth Farms, Jack Buck Farms, Caley Farms and Hay Farming. This provided a combined 2008-15 farmed area of 34,235 ha of which 2,446 ha was vining peas (Table 7) in a variety of rotational contexts (Appendix 5). In March (20-03-17) a follow-up meeting was held with these growers to feedback the LandIS derived outputs [Machinery Work Days (MWD); 'Structure susceptible to Topsoil Slaking (StoTS); Structure susceptible to Compaction (StoC); Potential of soil for Natural Structural Regeneration (NR)] and discuss the initial agri-informatics analysis of the Worth Farms dataset.

A key outcome from this meeting was that Richard Fitzpatrick (HMC Peas Grower Group Manager) and Simon Day (Worth Farms) indicated that they were happy for a summary of the Worth Farms Case study [included in the 2016 Annual Report] to be provided to addition Grower Group managers (listed below) in order to bring in addition whole farm datasets. These were contacted with regards providing data to SMIS.

In addition, soil analytical data not held within Gatekeeper was obtained from Jack Buck Farms 2008-15) and Hay Farming (2010-15). This data was subsequently input manually by Mark White of PGRO and added to the SMIS database.

- Ian Watson: Stemgold (Farmcare)
- Matthew Haywood: Swaythorpe Growers (Yorkshire Peas)
- [1]Andrew Lenson: Velcourt
- Stephen Francis (West Fen Peas)
- [2]Jim Hayes (Beeswax)

[1]Andrew Lenson has now left Velcourt and been replaced by Adrian Whitehead) [2]PGRO main contact at Beeswax is Edward Ford.

As a result of speculative e-mails, Matthew Haywood provided the contact details of three growers who agreed to provide whole farm Gatekeeper data to SMIS. Data collection meetings were held on 9th May 2017 with Charlie Parker (General Manager (Arable)) of JSR Farming and Andrew Johnston of Albanwise Yorkshire. In addition, Adam Horshield (Assistant Farm Manager) from Albanwise Norfolk was also present and subsequently provided significant data to SMIS (Table 6).

The purpose of these meetings was to

- Introduce the aims and objectives of the AHDB-Horticulture SMIS Project
- Elicit detailed expert knowledge and experiential evidence regarding the key soil management challenges of their grower groups and soil management options that have been adopted to address these challenges (with a specific focus on peas).

Key soil management issues identified by the Yorkshire based growers were as follows:

- Limited compaction issues except on tramlines and headlands
- High pH (circa high 7s to 8.0): Causes immobilisation of P, Zn, Fe, Mn and B. This is mitigated through application of trace element foliar applications.
- For JSR and Albanwise Yorkshire, Footrot was not an issue due to stoney nature of soils.

For Albanwise (Norfolk) key soil management issues included

- Undertaking sub-soiling operations at appropriate soil moisture conditions to optimise sub-soiling efficiency
- Experiential evidence of a link between soil compaction and Foot Rot
- Topsoil slaking/slumping and capping prior to and post emergence
- Avoiding growing peas following late harvest crops such as sugarbeet to avoid loss in yield associated with legacy-compaction.
- Key soil borne disease on peas is Foot Rot

- The need to extend rotations to 8 years to avoid build-up of key soil borne diseases impacting on yields (current pea rotation is 5-6 years).

These growers provided an additional farmed area for 2010-15 of 27,064 ha of which 1,517 ha was vining peas (Table 7).

In addition, the pea grower dataset contains 90% of the UK production area of celeriac with a total area of 441 ha between, 2010-15. Further, the dataset includes 664 and 174 ha of drilled onions and onion sets. Limited areas of chicory, fennel, daffodils, salads and French beans are also included (Table 8). To date with the exception of celeriac and vining peas, there is insufficient data to apply datamining and informatics approaches to evaluate soil management factors affecting yield and yield quality. Consequently, additional steps have been undertaken to source further horticultural datasets (Section 3.1.4.2). The results of applying agri-informatics approaches to the pea vining dataset will be presented at the Vining Pea conference on November 21st 2017.

It is important to note that additional intrinsic site factor and outcome datasets have been linked to the grower Gatekeeper data sets (Table 6). Foot Rot Index data has been input by Mark White who has also input farm soil analysis data (> 5 years worth of data from multiple fields) not on Gatekeeper for JBF and HF.

Table 6. Additional datasets linked to SMIS vining pea Gatekeeper datasets

SMIS Requirements Analysis	Data	WF	HF	JBF	CF	AY-LM	AY-R	AN	JSR
Intrinsic site factors/properties	[1]Foot Rot Index	✓	✓	✓	✓	R^1	R^1	R^2	R^1
Intrinsic site factors: LandIS derived outputs	MWD, StoTS, StoC, NR	✓	✓	✓	✓	✓	✓	✓	✓
Outcomes	[2]Pea yield & yield quality	✓	✓	✓	✓	R^1	R^1	✓	R^1

WF = Worth Farms; HF = Hay Farming; JBF = Jack Buck Farms; CF = Caley Farms; AY-LM = Albanwise Yorkshire (Low Mowthorpe); AY-R = Albanwise Yorkshire (Routh); AN = Albanwise Norfolk; JSR = JSR Ltd. R^1 = Data requested from Matthew Haywood; R^2 = Data requested from Russell Caulfield.

Table 7. Summary of current Pea Grower dataset indicating area (ha) of vining peas per annum.

	Year							
Grower	2008	2009	2010	2011	2012	2013	2014	2015
AY-LM	/	/	/	850 (0)	747 (0)	1246 (0)	825 (0)	1097 (**35**)
AY-R	/	/	/	966 (0)	699 (0)	1509 (0)	753 (0)	967 (0)
AN	/	/	/	1074 (**122**)	1138 (**108**)	890 (**92**)	1157 (**180**)	970 (**175**)
[1]JBF	434 (**28**)	480 (**45**)	520 (**43**)	482 (**38**)	494 (**44**)	690 (**44**)	486 (**38**)	340 (**41**)
[1]CF	785 (0)	986 (0)	1379 (**36**)	1516 (**40**)	1521 (0)	1875 (0)	2403 (0)	2607 (**67**)
[1]HF	/	/	666 (**100**)	886 (0)	728 (**58**)	1027 (**168**)	1037 (**105**)	1216 (**120**)
[1]WF	/	/	1752 (**304**)	1638 (**256**)	1875 (**238**)	1965 (**238**)	2144 (**255**)	1572 (**140**)
JSR	/	/	/	/	2159 (0)	3068 (**240**)	3386 (**300**)	3563 (**265**)
Total area (ha)	1219	1466	4317	7412	9961	12270	12191	11663
Vining peas (ha)	28	45	483	456	448	782	878	843

[1]WF = Worth Farms; [1]HF = Hay Farming; [1]JBF = Jack Buck Farms; [1]CF = Caley Farms; AY-LM = Albanwise Yorkshire (Low Mowthorpe); AY-R = Albanwise Yorkshire (Routh); AN = Albanwise Norfolk; JSR = JSR Ltd. [1]Whole farm data collected in 2016.

Table 8. Non-vining pea horticultural crops associated with the pea grower Gatekeeper datasets.

	Year											
Crop	2010		2011		2012		2013		2014		2015	
	Area (ha)	% of area	Area (ha)	% of area	Area (ha)	% of area	Area (ha)	% of area	Area (ha)	% of area	Area (ha)	% of area
Celeriac	58	8.6	60	22.3	78	36.6	59	25.1	87	20.5	99	15.7
Chicory	13	1.9	24	9.1	9	4.3	/	/	/	/	/	/
Fennel	22	3.3	17	6.4	16	7.6	22	9.3	13	3.0	19	3.0
Onions (Set)	31	4.6	14	5.2	32	15.0	31	13.3	33	7.6	33	5.2
Onions (Drilled)	66	9.8	153	57.0	71	33.3	113	48.0	120	28.2	141	22.4
Daffodils	/	/	/	/	6.6	3.1	10	4.2	/	/	/	/
Beans Dried Spring	/	/	/	/	60	28.3	/	/	76	17.7	337	53.3
Beans French	/	/	/	/	69	32.3	/	/	79	18.5	/	/
Salads	/	/	/	/	/	/	/	/	19	4.4	/	/
Squashes	/	/	/	/	/	/	/	/	/	/	3	0.5
Total area of horticultural crops (ha)	189	/	268	/	212	/	236	/	426	/	633	/

3.1.4.2. Horticultural crop data pipeline

a) Carrot Growers

The SMIS project aims and objectives were presented at the BCGA meeting on 29th of June 2017. At this meeting, Martin Evans was supportive and introduced Rob Simmons to a number of carrot growers who indicated that they would be willing to provide data to SMIS (Table 9).

To date, a partial Gatekeeper data set (2008-16, >1300 fields and 9886 ha) has been obtained from Ian Holmes at Strawsons with the remaining data expected by the end of November 2017. In addition, a date has been set with Jackie Seddon to obtain data from Tompsett Growers Ltd (Table 9).

During data collection an interview was conducted with Ian Holmes to gain insight into key soil management issues and how these are currently addressed. These were as follows:

- Cavity spot: This is a priority for product quality. SL567A applied at 2nd true leaf stage as a single application to all strawed crops.
- Capping: Impacts on seedling emergence, stand establishment and runoff. Use irrigation if available to prevent crust formation by keeping the soil surface moist. If irrigation not available and crust formed, a 'cap-buster' is used.
- Wind blow: Can lead to abrasive damage of seedling and loss of stand is addressed by plant barley rows. Strawsons have also sprayed Emupol SCC (water-based dispersion of acrylic copolymer) which is promoted by Hutchinsons for wind erosion control. Recommended application rate is 167kg of Emupol SCC mixed with 1200kg of water per ha.
- Runoff and erosion: Typically managed by trying to plant across the contour. Can also use a Dammer-Dyker.
- FLN: Use granular nematicide Vidate incorporated into seed bed.
- Sclerontia: Soil borne disease that produces spores that affect the canopy. This can also cause crown rot in carrots which is not manifest until packed and with customer. Current practice is to spray pre-emptive fungicide (6-9 times per crop). First application at 60% canopy so that product can penetrate canopy.

Table 9. Carrot grower contacts.

Grower Group	Contact/s	Confirmed data provision (Y/N)	Data Received (Y/N)	Source and date of contact

Tompsett Growers Ltd	Ian Hall & Jackie Seddon	Y	Collection date 13-10-17	BCGA Meeting (29-06-17)
Sherwood Produce	Philip Lilley	Y	N	
Strawsons	Ian Holmes	Y	Y*	
Kettle Produce	Euan Alexander	N	N	
Hardstaffs of Lindby	Michael Hardstaff	N	N	Provided by Martin Evans (05-09-17)
Poskitts	James Bramley	N	N	
Huntapac	Ben Madarasi	Y	N	BCGA Variety Day (05-10-17)
James Foskett Farms	Mike Shapland	Y	Collection date 13-10-17	Onion R&D Meeting (31-10-17)

*Partial dataset received remainder to be collected in November 2017.

In addition, at the June BCGA meeting in Martin Evans highlighted a previous AHDB funded project entitled 'Incidence of cavity spot in commercial carrots' (FV373 and FV373a).

Cavity spot is considered a high priority under the BCGA Research and Knowledge Exchange – Industry Priorities (2015-18). It is a serious and recurring disease of commercial carrots in the UK, which is largely unpredictable. Yield losses through cavity spot associated defects can be up to 40%. Current control systems rely on a single soil applied fungicide SL567A which is only partially successful.

Under FV373 and FV373a, during 2010-13 ninety commercial carrot production sites (30 per annum) provided by members of the BCGA and representative of the main carrot production areas of England and Scotland were monitored for total water input (precipitation and irrigation), soil moisture and soil temperature. At each site the incidence and severity of cavity spot disease was established by sampling prior to harvest and relationships were sought between the recorded site conditions and the incidence of disease. In addition, an extremely limited suite of soil analyses was undertaken at each site (Figure 9).

The outcomes of this study were inconclusive. Water input had some influence on disease development but it was not possible to say with any confidence what amount or time period is important. The authors state that the time around the onset of the main period of root expansion appears to be important but it was not impossible to predict with any confidence the incidence of disease from excess water at that time (AHDB, 2013 FV373 Final Report). In addition, soil temperature and soil saturation do not seem do have any relation to disease presence or absence.

A real opportunity exists to optimise the use of LandIS derived outputs and participating grower data (Figure 9) to augment the data obtained from FV373 and use agri-informatics data mining approaches to explore soil management drivers of cavity spot incidence and severity. Such insights would have a direct commercial benefit to UK carrot growers.

Consequently, the raw data from the FV373 project has been obtained and input to SMIS. In addition, current carrot grower data collection activities have focused on those growers who participated in the FV373 and FV373a project. In this regard, an e-mail was circulated via Coral Russel to BCGA members to request their participation. In addition, at the BCGA Variety Day held at Tomsett Growers (5th October 2017), Rob Simmons was introduced by Martin Evans to Ben Madarasi of Huntapac (Table 9). Huntapac provided a number of carrot production sites for the FV373 project.

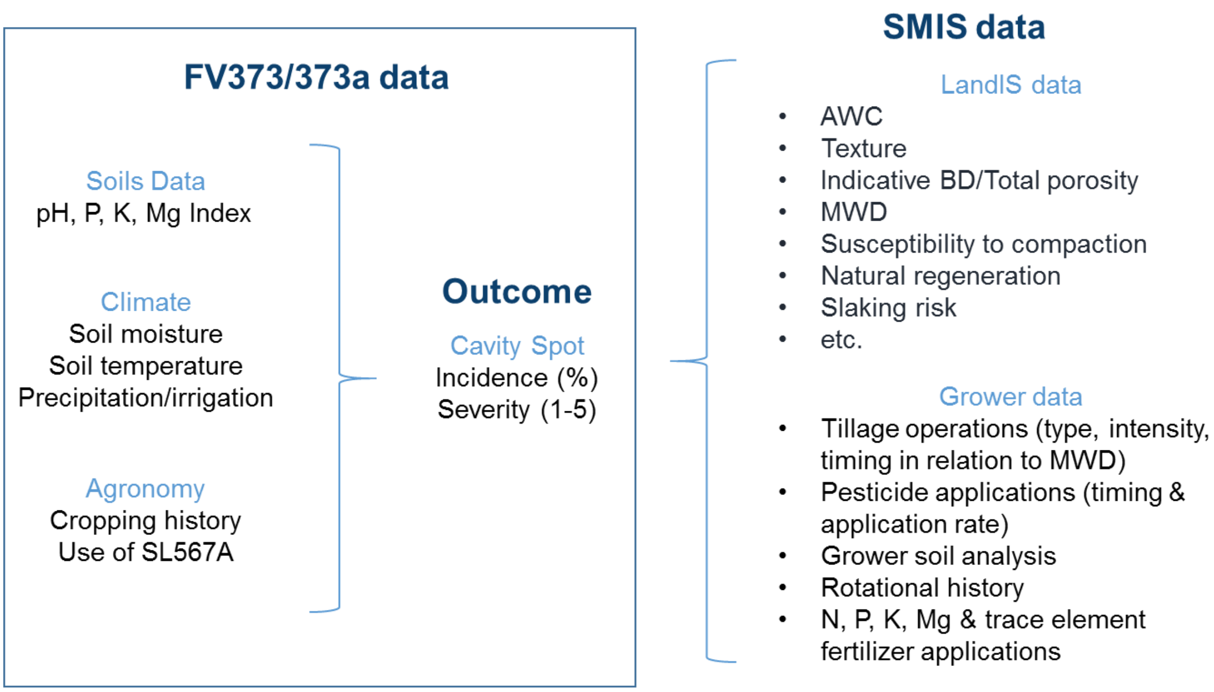


Figure 9. Opportunity to combine FV373/373a and LandIS and grower data to critically evaluate soil management practices affecting cavity spot incidence and severity

3.1.4.3. Growers associated with the CP107d Precision Farming project Soil Structure Survey

The soils data has been received from the CP107d Precision Farming project Soil Structure Survey. This data has been incorporated within SMIS. This dataset provides key soil structural metrics for 80 fields across a range of horticultural crops [Apples x 6, Asparagus x 6, Blackcurrants x 6, Brassicas x 16, Leeks x 7, Lettuce x 11, Onions x 5, Carrots/Parsnips x 10, Vining peas x 3, Raspberries x 4 and cut flowers x 6]. These metrics include bulk density, penetrative resistance, VSA, VESS, earthworm counts and presence/severity of a tillage pan. Basic soil characteristics including pH, texture, extractable P, K and Mg, P, K and Mg Indices, soil organic matter via loss on ignition (LOI), soil organic carbon and Total-N are also recorded.

Significant value can be added to this dataset if the soil metrics measured can be critically evaluated against field operations that have been undertaken on the fields sampled across their full rotational context. In this regard an e-mail was circulated via Paul Newell Price to growers that participated in the CP107d soil structural survey. To date two growers have responded namely Adrian Baker (Onions) from Parrish Farms and Stephen Barnes (Asparagus). Dates for data collection have been set for late November early December 2017. The e-mail will be re-sent to elicit further responses.

3.1.4.4. Grower Associations

To facilitate data download from Gatekeeper, a GK Tutorial document was produced (Appendix 10. SMIS USE CASE DOCUMENTATION V1.0

Revision History

Version	Description	Date	Author
1.0	Original SMIS use case report. "Established query" (see section 0) use cases will be updated as their viability is tested and their interfaces are implemented.	9/10/2017	Tomasz Kurowski

Abbreviations

API Application Programming Interface

NG National Grid

OS Ordnance Survey

REST Representational State Transfer

SMIS Soil Management Information System

1. INTRODUCTION

1.1. Purpose

The following document describes the 'use cases' defined for the SMIS Analytics Toolkit software developed as part of the SMIS project (AHDB CP107D). The document's primary purpose is to inform the end user about the defined uses of the system and the interaction flows required to achieve particular goals within its scope. On the developer side, the document will also serve as an implementation guide for the Analytics Toolkit, in particular for the design, development and installation of its front-end interface.

Additionally, as the inclusion of some use cases (see Section 0) will depend on validation of their viability during the remaining duration of the project, future versions of this document will effectively catalogue any additional use cases which are found to be viable and will therefore be included in the SMIS Analytics Toolkit.

1.2. Scope

This document describes the SMIS use cases without specifying implementation details beyond the overall flow of actor interactions the system is designed to allow. In particular, this means that both the back-end database, analytics and API functionalities are not discussed, and front-end interface details are not presented. Data gathering and manual curation activities undertaken within the scope of the SMIS project are also not discussed.

However, as the delineation of user roles (i.e. Operator and Administrator actors), as well as the classification of use cases used in this document result directly from certain system design and implementation decisions, a brief overview of the SMIS software design is provided in Section 2.

Additionally, a certain class of use case, the Established Query (described in Section 0) effectively consists of special cases of a more general use case, each intended to be used with its own custom interface and a more narrowly defined set of goals and user inputs. The list of Established Queries provided in the current version of this document is not intended to be exhaustive. These use cases and their associated interfaces will continue to be added to during the duration of the project as the viability (which is strongly dependent on gathered data) of various queries is verified and their usefulness validated. The document will be updated to include them as development continues.

2. OVERVIEW

The SMIS Analytics Toolkit software system can be considered to consist of several subsystems:

Parsing suite – a set of command-line tools developed for the purpose of importing external datasets (e.g. Grower Data, Experimental / Research Project Data, Literature Data) into the SMIS database, capturing their internal relationships and ensuring their integrity. This component interfaces with the database back-end directly and is the primary route of interacting with the system for Administrator actors.

Database back-end[3] – a non-relational database responsible for storing the parsed data gathered within the scope of the SMIS project, as well as the results of analyses conducted by the Analytics back-end. This will be available to the front-end and Analytics back-end through a REST API.

Analytics back-end – a set of scientific computing tools developed for generating analyses and visualisations for the SMIS system. Other than relatively trivial analyses like summarising available data, this subsystem primarily focuses on machine learning algorithms designed to identify relationships within the gathered datasets or their subsets. This is available to the front-end through a REST API.

Web application front-end – a web application which provides the primary interface for the delivery of the main SMIS use cases. This accepts user input and accesses the Database back-end and Analytics back-end through a REST API in order to present (or generate) results of user queries. This is the primary route for interacting with the system for Operator actors, and as such it represents the main SMIS Analytics Toolkit user interface in general.

2.1. Actors

2.1.1. Operator

The ‘Operator’ actor is the primary user role defined for the SMIS Analytics Toolkit. The role represents users whose primary goal is the exploration and extraction of useful information from the SMIS database. This can consist of browsing and searching through the raw (and curated) datasets, visualising their contents, exploring “rule bases” which represent the information derived from the stored datasets using machine learning, constructing queries which can be used to generate more specific rule bases, or using pre-defined “Established Queries” to explore problems which have been identified as being both useful and possible to address, based on the data collected during the course of the SMIS project.

2.1.2. Administrator

The ‘Administrator’ actor is a user role responsible for populating the SMIS database. During the development stage, this role can therefore be fully identified as the developer. However, the data parsing suite is intended to be generic and well-documented, capable of being used to update and expand the SMIS database beyond the scope of initial development, if further data is acquired. At the same time, the wholly separate interface (command-line scripts on the SMIS server) and very different input and server access requirements clearly delineate the role of an Administrator as separate from a “normal” user or Operator.

It should be noted that the SMIS database is generally intended for periodic, rather than continuous, updates. This is because the addition of new data may invalidate the results of previous analyses, requiring them to be re-computed. This means that the Administrator actor is intended to be active relatively infrequently.

An additional point is that as the SMIS Analytics Toolkit has been designed with the assumption that it will be used internally within an organisation rather than made directly available to the public, certain functionalities which would normally be assumed to depend on an “Administrator” role (such as user account management or access control), are not part of the design and are therefore not covered by this document.

2.2. Types of use cases

Three types of use cases have been defined. The first two (Exploratory and Analytical use cases) overlap to an extent, due to being available through a common Web interface and used by the same actor.

2.2.1. Exploratory

Exploratory use cases are ones which involve the use of the SMIS Analytics Toolkit Web application to browse raw and curated datasets stored in the SMIS database. They do not lead to the generation or storage of new information and do not involve the use of the Analytics back-end.

2.2.2. Analytical

Analytical use cases are ones which involve the use of the SMIS Analytics Toolkit Web application to view summaries and visualisations of datasets stored in the SMIS database, as well as to create queries which result in the creation of novel rule bases. These use cases employ the Analytics back-end to generate summaries, visualisations and machine learning models and store this new data in the database. The primary purpose of storing the results is to avoid the need to re-compute results for repeated queries, allowing for easy exploration of previously generated rule bases.

2.2.3. Administrative

Administrative use cases are ones which involve the parsing of novel data, resulting in either the addition of whole new datasets, or appending new data to existing ones. These cases involve the use of the SMIS parsing suite by an Administrator.

3. USE CASES

3.1. Browse and export grower data

Primary Actor: Operator

Type: Exploratory, Analytical (extensions)

Description: The Operator browses the SMIS Grower datasets and filters them by an arbitrary selection of the available data fields. The contents of the database for a selected set of conditions can be inspected and visualised, allowing the Operator to assess the size and potential usefulness of the gathered data for a given set of conditions. The visualisations and raw data can be exported.

Main success scenario:

The Operator accesses the SMIS Analytics Toolkit application through a Web browser.

The Operator navigates to the Grower data view.

A paginated table containing grower data is displayed.

The Operator selects (for qualitative fields) or types search terms for each column of interest.

Database contents are filtered and displayed in a new paginated table.

The Operator inspects the filtered data.

(Optional) The Operator presses a button to export the filtered data as a spreadsheet and the generated visualisations (if any) as a PDF report.

Extensions:

4a. The Operator can hide extraneous or empty table columns for clarity.

6a. (**Analytical**) Timeline view.

6a1. The Operator selects the Timeline view option.

6a2. A paginated list of farm fields and associated scrollable timelines of filtered field operations and applications are displayed in a rotational context. Field operations are coloured depending on an assessment as to whether they were undertaken in or outside of LandIS generated workability days.

6b. (**Analytical**) Summary view.

6b1. The Operator selects the Summary view option.

6b2. A summary report describing the filtered data is displayed, including the dates for which data is available, soil types and crop hectarages (including diagrams showing the relative hectarages of various crops and varieties).

6c. By selecting a data row, the Operator may navigate to a **Rule Base view** (see Section 0.) filtered to display rule bases which make use of this data point.

3.2. Browse and export experimental data

Primary Actor: Operator

Type: Exploratory

Description: The Operator browses the SMIS Experimental datasets, selecting the specific Experimental datasets to be viewed and inspecting the collected data in a tabular form, with the option to filter it by an arbitrary list of conditions and, optionally, export any filtered data subset.

Main success scenario:

The Operator accesses the SMIS Analytics Toolkit application through a Web browser.

The Operator navigates to the Experimental data view.

A list of available Experimental datasets is displayed along with short summaries of the size and content of each.

The Operator selects an Experimental dataset to view.

A paginated table containing Experimental data is displayed.

The Operator selects (for qualitative fields) or types search terms for each column of interest.

Database contents are filtered and displayed in a new paginated table.

The Operator inspects the filtered data.

(Optional) The Operator presses a button to export the filtered data as a spreadsheet.

Extensions:

4a. Alternatively, the Operator may navigate to a **Rule Base view** (see section 0.) filtered to display rule bases which make use of this experimental data set.

5a. The Operator can hide extraneous or empty table columns for clarity.

9a. By selecting a data row, the Operator may navigate to a **Rule Base view** (see section 0.) filtered to display rule bases which make use of this data point.

3.3. Browse and export literature data

Primary Actor: Operator

Type: Exploratory

Description: The Operator browses the SMIS Literature datasets, viewing and inspecting the manually curated data summaries in a tabular form, with the option to filter them by an arbitrary list of conditions and, optionally, export any filtered data subset.

Main success scenario:

The Operator accesses the SMIS Analytics Toolkit application through a Web browser.

The Operator navigates to the Literature data view.

A paginated table containing manually curated summaries of Literature data is displayed.

The Operator selects/enters search terms for each column of interest.

Database contents are filtered and displayed in a new paginated table.

The Operator inspects the filtered data.

(Optional) The Operator presses a button to export the filtered data as a spreadsheet.

Extensions:

4a. The Operator can hide extraneous or empty table columns for clarity.

6a. By selecting a literature entry, the Operator may navigate to a **Rule Base view** (see section 0.) filtered to display rule bases which make use of this literature entry.

3.4. View rule bases

Primary Actor: Operator

Type: Exploratory

Description: The Operator can view the soil management and yield “rule bases” derived from the SMIS database using machine learning analyses. These rule bases can be viewed either in tabular form, as lists of issues vs. causes/solutions (for soil management rule bases) or yield vs. factors-affecting-yield (for yield rule bases) pairs along with their associated weights, or in graph form, rendering the same pairwise relationships in a graphical manner.

Main success scenario:

The Operator accesses the SMIS Analytics Toolkit application through a Web browser.

The Operator navigates to the Rule Base view.

The Operator chooses to view either Soil Management Rule Bases or Yield Rule Bases.

A paginated list of available Soil Management / Yield rule bases is displayed alongside queries used to generate them. Rule bases still in the process of being generated are greyed out and display a progress bar with an estimate of time to completion.

The Operator selects a rule base to view.

A graphical view of the rule base is displayed. This consists of a set of nodes and edges connecting them. The nodes represent soil management issues, causes and solutions (for soil management rule bases) or crop yield and factors affecting crop yield (for yield rule bases). Edges connecting the nodes represent relationships identified based on the SMIS database contents according to a specific query. The thickness of edges depends on the strength and the degree of confidence of the relationship (weight).

The Operator inspects specific relationships by clicking on edges which display the list of Experimental, Grower and Literature evidence for the relationship along with their associated weights in absolute and relative terms.

(Optional) The Operator presses a button to export the entire rule base (or a selected set of nodes/edges) in tabular form as a spreadsheet.

Extensions:

6a. Tabular view.

6a1. The Operator selects the Tabular view option.

6a2. A paginated table containing a list of issues vs. causes/solutions (for soil management rule bases) or yield vs. factor-affecting-yield (for yield rule bases) pairs along with their associated weights is displayed.

6a3. The Operator types in search terms for each column of interest in the table.

6a4. The rule base contents are filtered and displayed in a new paginated table.

7a. By selecting a piece of Experimental, Grower or Literature evidence from the list, the Operator may navigate to their respective database browse views (see sections 3.1, 0, and 0), filtered to display the selected entry.

3.5. Query rule base

Primary Actor: Operator

Type: Analytical

Description: The Operator constructs a query used to select data used by a machine learning algorithm to generate a novel, specific rule base.

Main success scenario:

The Operator accesses the SMIS Analytics Toolkit application through a Web browser.

The Operator navigates to the Rule Base view.

The Operator chooses to view either Soil Management Rule Bases or Yield Rule Bases.

The Operator constructs a query by selecting available categories (data fields from Grower and Experimental data and curated Literature keywords), optionally combining them with a manually entered search term. A query can contain an arbitrary number of such category/search term pairs.

The Operator selects the "Run Query" option.

Database entries are filtered according to the query and the Analytics back-end begins the process of generating a rule base based on the filtered data.

The Operator is redirected to the rule base view (see section 0), with the new query and its progress bar highlighted.

Extensions:

6a. If no data stored by SMIS matches the entered query the Operator is informed of this by a pop-up message and the query is aborted.

6b. If a rule base related to the same or equivalent query already exists in the database, the Operator is informed of this by a pop-up message and navigates directly to a view of that rule base (see section 0).

3.6. Established queries

Established Queries can be considered "special cases" of rule base queries as used in Section 0. They rely on the same database and Analytics capabilities, but instead of allowing for the construction of arbitrary queries, they have their particular input and output interfaces, each of which are tailored to only one very specific query, selected during the SMIS project based on their potential usefulness and the confirmed availability of necessary data. One example is the assessment of soil compaction risk.

Assess soil compaction risk: an example of an established query

Primary Actor: Operator

Type: Analytical

Description: The Operator assesses the soil compaction risk for a specific field (identified by NG code / OS Map Sheet) at a particular range of dates, optionally associated with a specific crop.

Main success scenario:

The Operator accesses the SMIS Analytics Toolkit application through a Web browser.

The Operator navigates to the Established Queries view.

The Operator selects the Assess Soil Compaction Risk query.

The Operator enters the NG Code and Map Sheet for the field of interest.

The low-resolution LandIS data stored by SMIS is accessed to provide the most likely soil type for the field. This is displayed in a drop-down menu and can be changed manually.

The Operator enters a start and end date for the Soil Compaction Risk assessment and (optionally) selects the crop of interest from a drop-down menu.

A colour-coded timeline of workability days based on soil type and weather data is displayed and inspected by the Operator.

Extensions:

7a. By selecting the assessment results, the Operator can navigate to the associated rule base (see Section 0) and view possible solutions to soil compaction, if any were identified.

3.7. Import grower data

Primary Actor: Administrator

Type: Administrative

Description: The Administrator imports standard Grower datasets (e.g. Gatekeeper XML / spreadsheet format) into the SMIS database.

Main success scenario:

The Administrator accesses the SMIS server through a command line interface and launches the Grower Data Import script.

The Administrator selects the option to import a new dataset and provides a path to a Grower data file.

The Grower data is imported on a row-by-row basis. Novel data fields or qualitative field values require confirmation, modification or rejection by the user (see the *SMIS Database Architecture Technical Documentation* for details).

A data import report is displayed and can be saved as a text file.

3.8. Extend grower data

Primary Actor: Administrator

Type: Administrative

Description: The Administrator imports tabular Grower data not stored in a standard format such as Gatekeeper XML. Such data needs to be manually associated with datasets already present in the SMIS datasets and serve to add extra fields not covered by Gatekeeper datasets imported earlier.

Main success scenario:

The Administrator accesses the SMIS server through a command line interface and launches the Grower Data Import script.

The Administrator selects the option to expand an existing dataset and provides a path to a Grower data file.

Data fields present among Grower datasets stored in the SMIS database, as well as those identified in the provided file, are listed.

The Administrator selects one or more pairs of data fields to be considered equivalents to be used for joining the new dataset with data present in the database.

The Grower data is imported on a row-by-row basis. Novel data fields or qualitative field values require confirmation, modification or rejection by the user (see the *SMIS Database Architecture Technical Documentation* for details).

A data import report is displayed and can be saved as a text file.

3.9. Import experimental data

Primary Actor: Administrator

Type: Administrative

Description: The Administrator imports novel Experimental datasets into the SMIS database using project-specific parsing modules grouped under a common interface.

Main success scenario:

The Administrator accesses the SMIS server through a command line interface and launches the Experimental Data Import script.

The Administrator provides a path to a data file.

The Experimental data is imported by a project-specific parser. Unexpected data fields or field values may require confirmation, modification or rejection by the user (see the *SMIS Database Architecture Technical Documentation* for details).

A data import report is displayed and can be saved as a text file.

3.10 Import literature data

Primary Actor: Administrator

Type: Administrative

Description: The Administrator imports spreadsheets containing manually curated Literature data gathered during the SMIS project.

Main success scenario:

The Administrator accesses the SMIS server through a command line interface and launches the Literature Data Import script.

The Administrator provides a path to a data file.

The Experimental data is imported on a row-by-row basis

A data import report is displayed and can be saved as a text file.

Appendix 11). With the assistance of Jayne Dyas of British Growers, this was distributed in August to the following Associations/Producer Groups. A response date of end-November was set to avoid putting pressure on growers during peak periods. The e-mail will be re-sent in the 1st week of November.

- Asparagus Growers Assoc. [89 members]
- Brassica Growers Assoc. [80 members]
- British Onion Producers [89 members]
- Leek Growers Assoc. [22 members]
- British Leafy Salads Outdoor Group [60 members]
- Baby Leaf Growers Assoc. [18 members]
- British Herb Growers Assoc. [22 members]
- Outdoor Cucurbit Growers Assoc. [13 members]

3.1.5. Documentation outlining data required of growers/farmers (e.g. format of data and purpose of the data gathering)

The protocol for extracting data from Gatekeeper and the purpose of this is reproduced in Appendix 10. SMIS USE CASE DOCUMENTATION V1.0

Revision History

Version	Description	Date	Author
1.0	Original SMIS use case report. "Established query" (see section 0) use cases will be updated as their viability is tested and their interfaces are implemented.	9/10/2017	Tomasz Kurowski

Abbreviations

API Application Programming Interface

NG National Grid

OS Ordnance Survey

REST Representational State Transfer

SMIS Soil Management Information System

2. INTRODUCTION

3.10. Purpose

The following document describes the ‘use cases’ defined for the SMIS Analytics Toolkit software developed as part of the SMIS project (AHDB CP107D). The document’s primary purpose is to inform the end user about the defined uses of the system and the interaction flows required to achieve particular goals within its scope. On the developer side, the document will also serve as an implementation guide for the Analytics Toolkit, in particular for the design, development and installation of its front-end interface.

Additionally, as the inclusion of some use cases (see Section 0) will depend on validation of their viability during the remaining duration of the project, future versions of this document will effectively catalogue any additional use cases which are found to be viable and will therefore be included in the SMIS Analytics Toolkit.

3.11. Scope

This document describes the SMIS use cases without specifying implementation details beyond the overall flow of actor interactions the system is designed to allow. In particular, this means that both the back-end database, analytics and API functionalities are not discussed, and front-end interface details are not presented. Data gathering and manual curation activities undertaken within the scope of the SMIS project are also not discussed.

However, as the delineation of user roles (i.e. Operator and Administrator actors), as well as the classification of use cases used in this document result directly from certain system design and implementation decisions, a brief overview of the SMIS software design is provided in Section 2.

Additionally, a certain class of use case, the Established Query (described in Section 0) effectively consists of special cases of a more general use case, each intended to be used with its own custom interface and a more narrowly defined set of goals and user inputs. The list of Established Queries provided in the current version of this document is not intended to be exhaustive. These use cases and their associated interfaces will continue to be added to during the duration of the project as the viability (which is strongly dependent on gathered data) of various queries is verified and their usefulness validated. The document will be updated to include them as development continues.

4. OVERVIEW

The SMIS Analytics Toolkit software system can be considered to consist of several subsystems:

Parsing suite – a set of command-line tools developed for the purpose of importing external datasets (e.g. Grower Data, Experimental / Research Project Data, Literature Data) into the SMIS database, capturing their internal relationships and ensuring their integrity. This component interfaces with the database back-end directly and is the primary route of interacting with the system for Administrator actors.

Database back-end[3] – a non-relational database responsible for storing the parsed data gathered within the scope of the SMIS project, as well as the results of analyses conducted by the Analytics back-end. This will be available to the front-end and Analytics back-end through a REST API.

Analytics back-end – a set of scientific computing tools developed for generating analyses and visualisations for the SMIS system. Other than relatively trivial analyses like summarising available data, this subsystem primarily focuses on machine learning algorithms designed to identify relationships within the gathered datasets or their subsets. This is available to the front-end through a REST API.

Web application front-end – a web application which provides the primary interface for the delivery of the main SMIS use cases. This accepts user input and accesses the Database back-end and Analytics back-end through a REST API in order to present (or generate) results of user queries. This is the primary route for interacting with the system for Operator actors, and as such it represents the main SMIS Analytics Toolkit user interface in general.

4.1. Actors

4.1.1. Operator

The ‘Operator’ actor is the primary user role defined for the SMIS Analytics Toolkit. The role represents users whose primary goal is the exploration and extraction of useful information from the SMIS database. This can consist of browsing and searching through the raw (and curated) datasets, visualising their contents, exploring “rule bases” which represent the information derived from the stored datasets using machine learning, constructing queries which can be used to generate more specific rule bases, or using pre-defined “Established Queries” to explore problems which have been identified as being both useful and possible to address, based on the data collected during the course of the SMIS project.

4.1.2. Administrator

The ‘Administrator’ actor is a user role responsible for populating the SMIS database. During the development stage, this role can therefore be fully identified as the developer. However, the data parsing suite is intended to be generic and well-documented, capable of being used to update and expand the SMIS database beyond the scope of initial development, if further data is acquired. At the same time, the wholly separate interface (command-line scripts on the SMIS server) and very different input and server access requirements clearly delineate the role of an Administrator as separate from a “normal” user or Operator.

It should be noted that the SMIS database is generally intended for periodic, rather than continuous, updates. This is because the addition of new data may invalidate the results of previous analyses, requiring them to be re-computed. This means that the Administrator actor is intended to be active relatively infrequently.

An additional point is that as the SMIS Analytics Toolkit has been designed with the assumption that it will be used internally within an organisation rather than made directly available to the public, certain functionalities which would normally be assumed to depend on an “Administrator” role (such as user account management or access control), are not part of the design and are therefore not covered by this document.

4.2. Types of use cases

Three types of use cases have been defined. The first two (Exploratory and Analytical use cases) overlap to an extent, due to being available through a common Web interface and used by the same actor.

4.2.1. Exploratory

Exploratory use cases are ones which involve the use of the SMIS Analytics Toolkit Web application to browse raw and curated datasets stored in the SMIS database. They do not lead to the generation or storage of new information and do not involve the use of the Analytics back-end.

4.2.2. Analytical

Analytical use cases are ones which involve the use of the SMIS Analytics Toolkit Web application to view summaries and visualisations of datasets stored in the SMIS database, as well as to create queries which result in the creation of novel rule bases. These use cases employ the Analytics back-end to generate summaries, visualisations and machine learning models and store this new data in the database. The primary purpose of storing the results is to avoid the need to re-compute results for repeated queries, allowing for easy exploration of previously generated rule bases.

4.2.3. Administrative

Administrative use cases are ones which involve the parsing of novel data, resulting in either the addition of whole new datasets, or appending new data to existing ones. These cases involve the use of the SMIS parsing suite by an Administrator.

5. USE CASES

5.1. Browse and export grower data

Primary Actor: Operator

Type: Exploratory, Analytical (extensions)

Description: The Operator browses the SMIS Grower datasets and filters them by an arbitrary selection of the available data fields. The contents of the database for a selected set of conditions can be inspected and visualised, allowing the Operator to assess the size and potential usefulness of the gathered data for a given set of conditions. The visualisations and raw data can be exported.

Main success scenario:

The Operator accesses the SMIS Analytics Toolkit application through a Web browser.

The Operator navigates to the Grower data view.

A paginated table containing grower data is displayed.

The Operator selects (for qualitative fields) or types search terms for each column of interest.

Database contents are filtered and displayed in a new paginated table.

The Operator inspects the filtered data.

(Optional) The Operator presses a button to export the filtered data as a spreadsheet and the generated visualisations (if any) as a PDF report.

Extensions:

4a. The Operator can hide extraneous or empty table columns for clarity.

6a. (**Analytical**) Timeline view.

6a1. The Operator selects the Timeline view option.

6a2. A paginated list of farm fields and associated scrollable timelines of filtered field operations and applications are displayed in a rotational context. Field operations are coloured depending on an assessment as to whether they were undertaken in or outside of LandIS generated workability days.

6b. (**Analytical**) Summary view.

6b1. The Operator selects the Summary view option.

6b2. A summary report describing the filtered data is displayed, including the dates for which data is available, soil types and crop hectarages (including diagrams showing the relative hectarages of various crops and varieties).

6c. By selecting a data row, the Operator may navigate to a **Rule Base view** (see Section 0.) filtered to display rule bases which make use of this data point.

5.2. Browse and export experimental data

Primary Actor: Operator

Type: Exploratory

Description: The Operator browses the SMIS Experimental datasets, selecting the specific Experimental datasets to be viewed and inspecting the collected data in a tabular form, with the option to filter it by an arbitrary list of conditions and, optionally, export any filtered data subset.

Main success scenario:

The Operator accesses the SMIS Analytics Toolkit application through a Web browser.

The Operator navigates to the Experimental data view.

A list of available Experimental datasets is displayed along with short summaries of the size and content of each.

The Operator selects an Experimental dataset to view.

A paginated table containing Experimental data is displayed.

The Operator selects (for qualitative fields) or types search terms for each column of interest.

Database contents are filtered and displayed in a new paginated table.

The Operator inspects the filtered data.

(Optional) The Operator presses a button to export the filtered data as a spreadsheet.

Extensions:

4a. Alternatively, the Operator may navigate to a **Rule Base view** (see section 0.) filtered to display rule bases which make use of this experimental data set.

5a. The Operator can hide extraneous or empty table columns for clarity.

9a. By selecting a data row, the Operator may navigate to a **Rule Base view** (see section 0.) filtered to display rule bases which make use of this data point.

5.3. Browse and export literature data

Primary Actor: Operator

Type: Exploratory

Description: The Operator browses the SMIS Literature datasets, viewing and inspecting the manually curated data summaries in a tabular form, with the option to filter them by an arbitrary list of conditions and, optionally, export any filtered data subset.

Main success scenario:

The Operator accesses the SMIS Analytics Toolkit application through a Web browser.

The Operator navigates to the Literature data view.

A paginated table containing manually curated summaries of Literature data is displayed.

The Operator selects/enters search terms for each column of interest.

Database contents are filtered and displayed in a new paginated table.

The Operator inspects the filtered data.

(Optional) The Operator presses a button to export the filtered data as a spreadsheet.

Extensions:

4a. The Operator can hide extraneous or empty table columns for clarity.

6a. By selecting a literature entry, the Operator may navigate to a **Rule Base view** (see section 0.) filtered to display rule bases which make use of this literature entry.

5.4. View rule bases

Primary Actor: Operator

Type: Exploratory

Description: The Operator can view the soil management and yield “rule bases” derived from the SMIS database using machine learning analyses. These rule bases can be viewed either in tabular form, as lists of issues vs. causes/solutions (for soil management rule bases) or yield vs. factors-affecting-yield (for yield rule bases) pairs along with their associated weights, or in graph form, rendering the same pairwise relationships in a graphical manner.

Main success scenario:

The Operator accesses the SMIS Analytics Toolkit application through a Web browser.

The Operator navigates to the Rule Base view.

The Operator chooses to view either Soil Management Rule Bases or Yield Rule Bases.

A paginated list of available Soil Management / Yield rule bases is displayed alongside queries used to generate them. Rule bases still in the process of being generated are greyed out and display a progress bar with an estimate of time to completion.

The Operator selects a rule base to view.

A graphical view of the rule base is displayed. This consists of a set of nodes and edges connecting them. The nodes represent soil management issues, causes and solutions (for soil management rule bases) or crop yield and factors affecting crop yield (for yield rule bases). Edges connecting the nodes represent relationships identified based on the SMIS database contents according to a specific query. The thickness of edges depends on the strength and the degree of confidence of the relationship (weight).

The Operator inspects specific relationships by clicking on edges which display the list of Experimental, Grower and Literature evidence for the relationship along with their associated weights in absolute and relative terms.

(Optional) The Operator presses a button to export the entire rule base (or a selected set of nodes/edges) in tabular form as a spreadsheet.

Extensions:

6a. Tabular view.

6a1. The Operator selects the Tabular view option.

6a2. A paginated table containing a list of issues vs. causes/solutions (for soil management rule bases) or yield vs. factor-affecting-yield (for yield rule bases) pairs along with their associated weights is displayed.

6a3. The Operator types in search terms for each column of interest in the table.

6a4. The rule base contents are filtered and displayed in a new paginated table.

7a. By selecting a piece of Experimental, Grower or Literature evidence from the list, the Operator may navigate to their respective database browse views (see sections 3.1, 0, and 0), filtered to display the selected entry.

5.5. Query rule base

Primary Actor: Operator

Type: Analytical

Description: The Operator constructs a query used to select data used by a machine learning algorithm to generate a novel, specific rule base.

Main success scenario:

The Operator accesses the SMIS Analytics Toolkit application through a Web browser.

The Operator navigates to the Rule Base view.

The Operator chooses to view either Soil Management Rule Bases or Yield Rule Bases.

The Operator constructs a query by selecting available categories (data fields from Grower and Experimental data and curated Literature keywords), optionally combining them with a manually entered search term. A query can contain an arbitrary number of such category/search term pairs.

The Operator selects the "Run Query" option.

Database entries are filtered according to the query and the Analytics back-end begins the process of generating a rule base based on the filtered data.

The Operator is redirected to the rule base view (see section 0), with the new query and its progress bar highlighted.

Extensions:

6a. If no data stored by SMIS matches the entered query the Operator is informed of this by a pop-up message and the query is aborted.

6b. If a rule base related to the same or equivalent query already exists in the database, the Operator is informed of this by a pop-up message and navigates directly to a view of that rule base (see section 0).

5.6. Established queries

Established Queries can be considered "special cases" of rule base queries as used in Section 0. They rely on the same database and Analytics capabilities, but instead of allowing for the construction of arbitrary queries, they have their particular input and output interfaces, each of which are tailored to only one very specific query, selected during the SMIS project based on their potential usefulness and the confirmed availability of necessary data. One example is the assessment of soil compaction risk.

Assess soil compaction risk: an example of an established query

Primary Actor: Operator

Type: Analytical

Description: The Operator assesses the soil compaction risk for a specific field (identified by NG code / OS Map Sheet) at a particular range of dates, optionally associated with a specific crop.

Main success scenario:

The Operator accesses the SMIS Analytics Toolkit application through a Web browser.

The Operator navigates to the Established Queries view.

The Operator selects the Assess Soil Compaction Risk query.

The Operator enters the NG Code and Map Sheet for the field of interest.

The low-resolution LandIS data stored by SMIS is accessed to provide the most likely soil type for the field. This is displayed in a drop-down menu and can be changed manually.

The Operator enters a start and end date for the Soil Compaction Risk assessment and (optionally) selects the crop of interest from a drop-down menu.

A colour-coded timeline of workability days based on soil type and weather data is displayed and inspected by the Operator.

Extensions:

7a. By selecting the assessment results, the Operator can navigate to the associated rule base (see Section 0) and view possible solutions to soil compaction, if any were identified.

5.7. Import grower data

Primary Actor: Administrator

Type: Administrative

Description: The Administrator imports standard Grower datasets (e.g. Gatekeeper XML / spreadsheet format) into the SMIS database.

Main success scenario:

The Administrator accesses the SMIS server through a command line interface and launches the Grower Data Import script.

The Administrator selects the option to import a new dataset and provides a path to a Grower data file.

The Grower data is imported on a row-by-row basis. Novel data fields or qualitative field values require confirmation, modification or rejection by the user (see the *SMIS Database Architecture Technical Documentation* for details).

A data import report is displayed and can be saved as a text file.

5.8. Extend grower data

Primary Actor: Administrator

Type: Administrative

Description: The Administrator imports tabular Grower data not stored in a standard format such as Gatekeeper XML. Such data needs to be manually associated with datasets already present in the SMIS datasets and serve to add extra fields not covered by Gatekeeper datasets imported earlier.

Main success scenario:

The Administrator accesses the SMIS server through a command line interface and launches the Grower Data Import script.

The Administrator selects the option to expand an existing dataset and provides a path to a Grower data file.

Data fields present among Grower datasets stored in the SMIS database, as well as those identified in the provided file, are listed.

The Administrator selects one or more pairs of data fields to be considered equivalents to be used for joining the new dataset with data present in the database.

The Grower data is imported on a row-by-row basis. Novel data fields or qualitative field values require confirmation, modification or rejection by the user (see the *SMIS Database Architecture Technical Documentation* for details).

A data import report is displayed and can be saved as a text file.

5.9. Import experimental data

Primary Actor: Administrator

Type: Administrative

Description: The Administrator imports novel Experimental datasets into the SMIS database using project-specific parsing modules grouped under a common interface.

Main success scenario:

The Administrator accesses the SMIS server through a command line interface and launches the Experimental Data Import script.

The Administrator provides a path to a data file.

The Experimental data is imported by a project-specific parser. Unexpected data fields or field values may require confirmation, modification or rejection by the user (see the *SMIS Database Architecture Technical Documentation* for details).

A data import report is displayed and can be saved as a text file.

3.10 Import literature data

Primary Actor: Administrator

Type: Administrative

Description: The Administrator imports spreadsheets containing manually curated Literature data gathered during the SMIS project.

Main success scenario:

The Administrator accesses the SMIS server through a command line interface and launches the Literature Data Import script.

The Administrator provides a path to a data file.

The Experimental data is imported on a row-by-row basis

A data import report is displayed and can be saved as a text file.

Appendix 11.

3.1.6. Muddy Boots

We have contacted Muddy Boots on several occasions but have had no response from them, either from the support/helpdesk staff, Farm Services Team or their business development / commercial team. We have good contacts with a Muddy Boots user (Mr John Chinn of Cobrey Farms, an asparagus and blueberry grower from Herefordshire) who has consented for us to use his data. Muddy Boots (Paul Thomas, Senior Business Development Manager) have suggested we obtain a manual data exchange file (e.g. Crop Walker) from Mr Chinn, but this won't meet our requirements in the medium to long term in developing a sustainable integration plan with the Muddy Boots system that will support future growers sharing their data via a consistent interface.

3.1.7. Integrating grower data with LandIS, RPA field boundaries and Met data

For the purposes of the SMIS application a 1km dataset has been derived to help assess the Gatekeeper and field information held in SMIS in terms whether the fields are likely to be vulnerable to compaction or topsoil slaking and whether it is able to recover naturally if inappropriate actions are performed. Also provided are the workability dates in a median year (plus a wet and dry quartile).

There are many different soil types likely in each 1km square each of which have distinct classifications. For each 1km square therefore the most likely class only is provided for each of the risk classes (compaction, topsoil slaking and natural regeneration). The most abundant workability class is then used to give the soil weighting which is then applied to the field capacity data to estimate the machinery work day period in a median, dry and wet quartile.

The 1km summary data (Table 10) will be provided without additional charge for the duration of the SMIS project. The on-going licencing of the data after this period depends on what arrangements are made for the on-going use of the SMIS application. It is recommended that the data is provided in an encrypted format that can only be accessed through the SMIS interface. This will mean no additional licencing will be required over and above that agreed for SMIS itself.

Table 10. LANDIS DATA FOR SMIS (1 Km)

Attribute	Description
MAPSHEET	1km ID used in Gatekeeper in field identification i.e. TF1234
WA	Most abundant workability assessment of soil series in 1km area

WA_PC	Proportion of 1km area covered by series in this workability assessment
S_to_TS	Most abundant topsoil slaking risk of soil series
TS_PC	Proportion of 1km area covered by series topsoil slaking class
S_to_C	Most abundant compaction risk of soil series
COMP_PC	Proportion of 1km area covered by series compaction class
NR	Most abundant natural regeneration risk of soil series
NR_PC	Proportion of 1km area covered by series natural regeneration class
FROM_WET	Date in autumn from which landwork may damage soil (in a wet quartile year)
FROM_MED	Date in autumn from which landwork may damage soil (in a median year)
FROM_DRY	Date in autumn from which landwork may damage soil (in a dry quartile year)
TO_DRY	Date in spring from which landwork is unlikely to damage soil (in a dry quartile year)
TO_MED	Date in spring from which landwork is unlikely to damage soil (in a median)
TO_WET	Date in spring from which landwork is unlikely to damage soil (in a wet quartile year)

To identify the regional climate for any site, a spreadsheet was generated for the regions in England and Wales shown in Figure 9 from regional MetOffice data describing whether each of the years between 2000 and 2016 were considered to be Dry, Wet or Median in Autumn and Spring.

Figure 10. Climate regions in UK

3.2. Building SMIS Backend

3.2.1. A database technical documentation developed including the supported dataset types

The SMIS Database Architecture Technical Documentation (Appendix 9) describes the database design decisions taken, the technical and practical considerations which guided them, and the intended uses of the database alongside the system context in which it will be used. The *upstream* context, that is the data parsing and validation processes involved in populating and organising the database based on the SMIS project data collection, was given particular focus, while the *downstream* context, that is how the data will be used by other subsystems, was described in more general terms.

3.2.2. Use Case scenario report describing data browsing and visualisation

The SMIS Use Case Documentation (Appendix 10) describes the intended uses of the SMIS Analytics Toolkit from the point of view of the end users, without discussing implementation details beyond basic, necessary concepts. User roles and interaction flows for the finished system were defined alongside a general discussion of the user interfaces involved in individual use cases.

Additionally, as the software system is under continued development, the technical documents are intended to undergo updates to reflect any changes and refinements in the design. This is of particular importance to the SMIS Use Case Documentation, which is explicitly intended to include more details on certain subsystems as their viability is validated during the course of the project.

Prototype applications following these specifications are in continued development according to the project schedule.

3.2.3. Completed Web interface final report (including data access, and browsing)

This is a technical report describing the data structure and the UI design. This includes data access and browsing, but not the data visualisation module, which will be implemented during the 3rd year of the project. This report is due end May 2018

3.3. SMIS UI and Analytics Toolkit

3.3.1. Technical documentation and user manual report completed

A technical report describing the data structure and the UI design will be submitted. This includes data access and browsing but not the data visualisation module, which will be implemented during the 3rd year of the project

Due end of July 2018

3.3.2. Full access to the developed SMIS system granted to client for final feedback to be received.

Due end of September 2018

3.3.3. Final SMIS implementation and documentation to be formally handed over to AHDB.

Due end of October 2018

3.4. Project Management

3.4.1. Contractual issues

Originally the contract ran from 1/06/15, but this was amended to 01/11/15, due to contract issues: paperwork was initially incorrect and for various reasons (e.g. JHI being incorrectly identified as a CoI rather than subcontractor), this was not resolved until October 2015. Contract signed on 28/10/2015. Project end date is 31/10/18. Following a project review in early 2017, a new contract between AHDB and Cranfield (with revised and more detailed milestones and deliverables) was signed on 24th July 2017, with the same start and end dates as before.

3.4.2. Revised risk register written and agreed

A revised risk register was agreed at the meeting on 31/03/17. It is reviewed at Quarterly meetings.

3.4.3. Quarterly updates to AHDB-Horticulture

These are on-going. Minutes and actions are available.

3.4.4. Monthly updates to AHDB by telephone

These are on-going. Minutes and actions are available.

4. DISCUSSION

The SMIS database has been expanded considerably over the past year, with over 581 items of literature with broad coverage of horticultural sectors being screened. Of these, only 84 were directly relevant to the scope of SMIS. The literature covers 20 broad soil management solutions with a focus towards generic 'management practice' (organic versus conventional) and rotation based solutions. The knowledge from the literature has been categorised to reflect certainty and confidence in these sources (including academic papers published in scientific, peer reviewed journals; conference proceedings / papers; research reports; the grey literature (e.g. articles on websites and in trade magazines)). The weight of evidence is reflected in whether the source is quantitative (based on empirical evidence from field work); qualitative (based on observations during a field-based experiment); or anecdotal (unreferenced statements). These categories reflect the degree of certainty and confidence associated with each and will be visualised within the SMIS User Interface, when this is developed in Year 3 of the project. This categorisation will inform the strength of relationships in the SMIS rules base (e.g. soil management practices and effects on yields, soil health etc.).

Better engagement with growers over the year has led to a number of datasets to be uploaded to the database. However, receiving data from growers is still a challenge. There are a number of reasons for this: heavy workloads throughout the year; limited time 'in the office' to access the data; concerns about the anonymity of the data; concerns about how clean and extensive the data might be – and whether it is any use to the project, especially if data is not recorded. Although the protocol for uploading data is self-explanatory (Appendix 11), many growers prefer to meet in person to go through their data records. This can be time consuming and is not sustainable in the long term (i.e. post project). It has not been possible to obtain data from Muddy Boots software, but Gatekeeper records, individuals' personal copies of field data and some paper records have been received, formatted and inputted to SMIS.

Separate to this, data-rich projects such as the FV5 cavity spot projects (1990 – 2007) have been enhanced with new soils data. These projects were found to be lacking in basic soils data, and with no significant differences observed in tested variables. With the use of geographic data collected in the reports, LandIS data such as that used within the SMIS project has been applied to consider new relationships between treatments.

Regarding the development of the SMIS system architecture, the set of software which together form a group of key deliverables of the Soil Management Information System (SMIS) project underwent a significant redesign during the course of the year. The changes included a reformulation of the software project structure and an explicit articulation of specific assumptions about how the completed system will operate, what technologies and frameworks it will depend upon, what its requirements will be, what types of Use Cases it will support, and what sort of results it will produce.

A real opportunity exists to optimise the use of LandIS derived outputs and participating grower data (Figure 9) to augment the data obtained from FV373 and other similar projects. The agri-informatics data mining approaches being developed in Year 3 will explore soil management drivers of cavity spot incidence and severity. Such insights would have a direct commercial benefit to UK carrot growers.

Knowledge exchange activities have been more frequent in Year 2, raising awareness of SMIS and its benefits to growers. The key objective has been to engage with growers that are willing to share their data. However, this is still a challenge for reasons stated above. As the benefits of SMIS become more tangible, it is hoped that more growers will input data, so developing and strengthening the rules base of SMIS.

4.1. Next steps

- Continue to source data / information / knowledge as input to SMIS database.
    - New literature and outputs from research projects will be scanned and reported to Quarterly meetings.
    - A data pipeline of growers has been established and members of the team will use the protocol to access the anonymised data from these growers.
- Integrate the knowledge identified in the literature review (in quantitative, qualitative and / or anecdotal form), the grower data (e.g. Gatekeeper records), findings from research projects and expert knowledge/ opinions (e.g. outputs from the Fuzzy Cognitive Mapping exercise) within the SMIS architecture.
- Develop the analytics toolkit to interrogate the database
- Develop the SMIS User Interface (front end) – technical documentation will include the specification of the system and user manual
- Devise case studies of soil management related queries that SMIS will address. These queries will be compared with the availability and accessibility of data (and associated) rules bases currently within the SMIS database. As the database within SMIS develops, it is envisaged that data mining techniques will provide useful insights to address AHDB Horticulture Panels' 2015-2018 priorities. Statistical interpretation of grower datasets within SMIS will provide a more scientific basis for guidance on a wider range of soil management issues (only soil compaction has been considered thus far). These issues might include (and are not limited to):
    - the selection and role of cover crops (GAEC Rule 4) in a range of rotational contexts to address key soil borne diseases affecting the horticultural sector through 'bio-fumigation',
    - enhancement of water and nutrient use efficiency through promotion of Arbuscular Mycorrhizal Fungal associations with direct impacts of yield and yield quality as well as reduced reliance on a diminishing range of chemical solutions.

    - soil structural improvement via ‘bio-drilling’ thus reducing reliance on costly and often ineffectual mechanical options
    - Sclerotinia: Vining Peas: Development of improved control strategy as disease becoming more frequent
    - Root diseases, including Fusarium Foot Rot [Utilizing PGRO Foot Rot Index data]: (An increasing problem with no chemicals available). An evaluation of cultural methods of suppression is required and the use of mustard bio fumigant cover crops.
    - Crop Nutrition: More information needed on P and K requirements.
- Continue to publicise and promote SMIS to interested parties, especially grower groups and associations such as the Field Veg Panel. Notice of upcoming meetings from AHDB staff will populate the table of events.
- Consider technical and commercial implications of where SMIS will reside post project

5. CONCLUSIONS

- The SMIS database has been built and comprises diverse sources of data, knowledge and information, including published literature, grower data and experimental data from other research projects, including others in CP107. Unique merging of grower data with LandIS data on soils and meteorology has been used to identify risks of compaction and structural degradation (e.g. slaking, surface capping), windows of opportunity for non-soil damaging field operations (workability days)
- Linkage with other GREAT soils projects has the potential to add significant value to the SMIS dataset, if the soil metrics measured can be critically evaluated against field operations that have been undertaken on the fields sampled across their full rotational context.
- Wider engagement with industry has taken place in 2017, extending contacts into brassica and carrots growers (identified in discussions) and to other horticultural groups. Better engagement with the industry had resulted in more datasets being included in the SMIS database, and clearer understanding of the purpose of SMIS and of the project outcomes for farmers and growers. We have received positive industry engagement, including their understanding of potential benefits and contributions to the development of the project.
- The SMIS system architecture underwent a significant redesign during the course of the year. The changes included a reformulation of the software project structure and an explicit articulation of specific assumptions about how the completed system will operate, what technologies and frameworks it will depend upon, what its requirements will be, what types of Use Cases it will support, and what sort of results it will produce.
- The queries to be run in SMIS are taking shape: one example is the opportunity to optimise the use of LandIS derived outputs and participating grower data (Figure 9) to augment the data obtained from FV373 and use agri-informatics data mining approaches to explore soil management drivers of cavity spot incidence and severity. Such insights would have a direct commercial benefit to UK carrot growers.

6. KNOWLEDGE AND TECHNOLOGY TRANSFER

This section describes the knowledge exchange activities carried out as part of the SMIS project. These activities were carried out by members of the project team in conjunction with Dr Lynda Deeks, NERC Horticulture Knowledge Exchange Fellow.

6.1. Attendance at industry events

Over the current life of the project, we have attended and participated at 18 horticultural meetings and have also hosted a stakeholder meeting at Cranfield University (June 2017). Figure 11 shows the sector specific distribution of events attended so far. Farming Groups events represent multiple horticultural sectors at a single event.

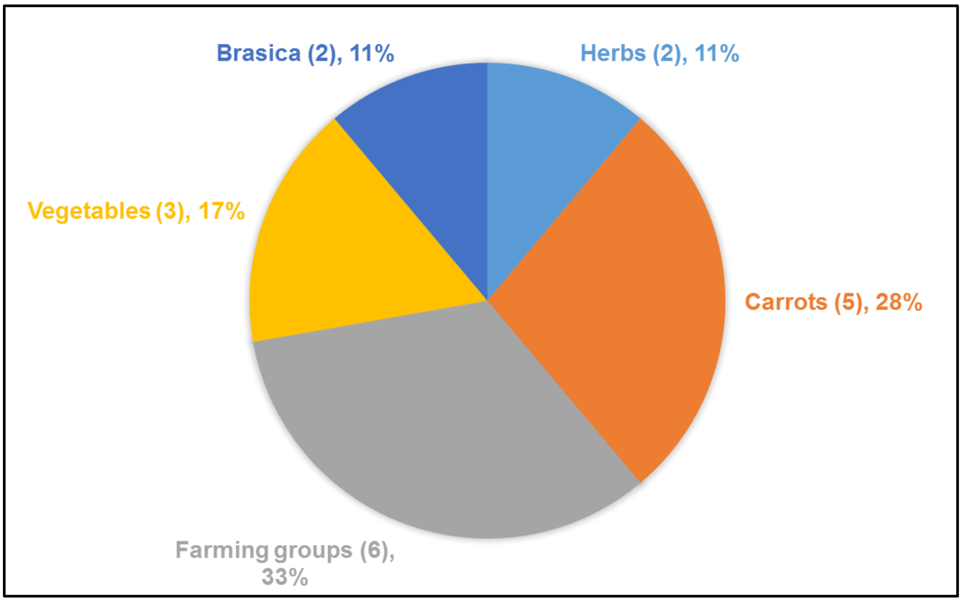


Figure 11. Horticultural events by sector attended, where the CP107d SMIS project was discussed

During Year 2 of the SMIS project specifically (2016-17), knowledge exchange activities have included attending and participating at 12 horticultural sector events, hosting a stakeholder meeting at Cranfield University (26th June 2017) and writing an article for the Grower Magazine (AHDB Grower 2017 – Oct/Nov, Great soils – CP107d SMIS, Big Data, Big Knowledge).

The events attended and presented at are detailed in Appendix 7. The objectives at these events were to promote the SMIS project to the horticultural community, to engage with individuals and/or companies who may be able to provide access to farm data (Gatekeeper, Muddy Boots^{tm} or other forms of farm records) and to learn from the horticultural community how they would wish to interact with SMIS.

Engagement with the horticultural community is helping to build a better picture of factors affecting soil management and therefore soil health. This includes economic drivers, factors affecting rotation decisions, current levels of data collection and utilisation, and likes and dislikes with regards to advice delivery. The community is cautious with regards to sharing data, but attendance at these events is helping to break down barriers as the value of SMIS becomes more apparent.

6.1.1. HMC Grower Group (March 2017)

The outcomes of the initial analyses undertaken on the Worth Farms data along with the LandIS derived maps for all the growers will be presented to the HMC Grower Group (Pers. Communication. Richard Fitzpatrick Grower Group Manager). It is expected that this will act as a catalyst for other growers to provide their Gatekeeper data.

6.1.2. AHDB Legume Panel (February 2017)

The outcomes of the initial analysis of Gatekeeper data (and ancillary supporting databases) will be presented to the AHDB Legume Panel in the form of a 2-page summary in early 2017 (Pers. Communication Becky Howard). It is anticipated that this will result in further data being obtained from additional grower groups. Further, in November 2017 a verbal presentation will be made to this panel (pers. comm., Becky Howard).

6.2. SMIS Stakeholder Workshop (June 2017)

The purpose of the workshop was to introduce the SMIS project and current progress to key stakeholders; to receive feedback on what and how SMIS can deliver to the industry; and demonstrate the potential benefits of SMIS to the industry. A detailed report on the Workshop can be found in Appendix 6. In summary, at first, there was some scepticism of the SMIS project from a few delegates, but the project team managed to reassure those present with our vision plus the project progress to date, including the grower data we have amassed and the tools being developed to interrogate the SMIS database. We also understood why Martin Evans and the Field Veg Panel in particular felt disengaged with the project. Rob Simmons met with Martin later in the week and Martin has put us in touch with a number of growers who we can approach for their farm data to feed into SMIS. Cathryn Lambourne has also offered to put us in touch with a number of grower associations and farmers groups to whom we can publicise the project and encourage them to be involved. (We already have a number of dates in the diary to speak to these groups). The break out sessions were lively with valuable contributions from all delegates. There was very positive feedback from industry representatives that we were undertaking innovative approaches to use ‘big data’ from growers to inform better soil management. Inevitably the future of SMIS after the end of the project (Nov 2018) was discussed, in terms of it being ‘owned’ by AHDB or by the growers / industry. This is an area for further discussions with the AHDB and Cranfield project teams.

6.3. The Grower article (Appendix 8)

The intention of the AHDB Grower magazine article was to promote the SMIS project to the wider horticultural community in order to promote interest in the benefits of sharing data. It was also intended, with its original publication timing to promote interest in the project to encourage participation at the stakeholder workshop. The delay in publication that was outside our control has meant that this article has not yet been as impactful as it was intended but it may still be beneficial to the project.

6.4. Future Knowledge Exchange activities

Planed attendance at future evens is also given in Appendix 7. Future events will be used to continue to promote the SMIS project and to promote the benefits of the power of shared data sets.

7. GLOSSARY

The project embodies a great many terms and concepts for which there needs to be a common reference and understanding. The following table provides working definitions for the common terms and concepts used in the development of SMIS. This will be updated as the project progresses.

Term	Definition
AHDB	Agriculture and Horticulture Development Board.
API	Application Programming Interface
Criteria	One of six metadata descriptions that can be completed to describe a given source of data. These comprise descriptions concerning place; time; characteristics; land use; operations; and outcomes. Typically this is recorded with an entry in a table for each item considered (e.g. an academic paper).
DBMS	Database Management System

e-Guide	Knowledge-Based System for presenting options, outcomes and best practices for soil management with relation to horticultural practices. A key project delivery.
HTTP(S)	Hypertext Transfer Protocol (Secure)
JSON	JavaScript Object Notation
LandIS	Land Information System (LandIS; http://www.landis.org.uk/) is a soils-focussed information system for England and Wales. http://www.cranfield.ac.uk/About/Cranfield/Themes/Agrifood LandIS, the "Land Information System", is a substantial environmental information system operated by Cranfield University, UK, designed to contain soil and soil-related information for England and Wales including spatial mapping of soils at a variety of scales, as well as corresponding soil property and agro-climatological data. LandIS is the largest system of its kind in Europe and is recognised by UK Government as the definitive source of national soils information. The Cranfield Soil and AgriFood Institute (CSAI), incorporating the National Soil Resources Institute (NSRI), is a centre within Cranfield University, and maintains this extensive geographic database of land-related data, covering England and Wales. Outlined below and within this site are a number of ways by which you can access this information.
MEAN	MongoDB, Express.js, AngularJS, and Node.js
NG	National Grid
OS	Ordnance Survey
Project	Project "Development of a Horticultural Soil Management Information System (SMIS). CP 107d/3110107425", sponsored by AHDB under the CP 107: Soils - Improved Sustainable Management for Horticultural Crops programme.

RDF	Resource Description Framework (https://www.w3.org/RDF/ and http://www.w3.org/TR/rdf11-concepts/), a means of recording semantic knowledge in computer-compatible form, using the concept of 'graphs', containing sets of subject-predicate-object triplets.
REST	Representational State Transfer
SMIS	Soil Management Information System. An information repository that contains a rule base, and supporting evidence from a range of sources. A key project delivery.
Soil Management Challenge	One of the soil management challenges identified by Rickson and Deeks (2013) that the SMIS sets out to address, identifying thematic areas of concern in the development and implementation of best practice guidelines for sustainable soil management. An example is 'soil compaction'.
SubVESS	Subsoil Visual Evaluation of Soil Structure
VESS	Visual Evaluation of Soil Structure
VSA	Visual Soil Assessment
W3C	The World Wide Web Consortium (https://www.w3.org), owners of the RDF schema.
XML	Extensible Markup Language

8. REFERENCES

References from the literature review are given in Appendix 2

9. APPENDICES

Appendix 1. SMIS project Gantt chart summarising Milestones and Deliverables

Appendix 2. Literature review report

1. Introduction

This report meets CP107d Development of a Horticultural Soil Management Information System (SMIS) deliverable 2.1 'Data collection and collation (literature review)'. Detailed within this report is the methodology used for the literature review, details of the reviewed literature (as related to the CP107a end user requirements analysis (Rickson and Deeks, 2013)), and the nature of the datasets available that will provide information and guidance to SMIS and its intended end users.

Please refer to the accompanying Excel spreadsheet 'Literature Review inventory' for details of the items that provide data for input into the SMIS database. The Excel file also includes 10 examples of the type of quantitative data to be inputted into SMIS and how this data will contribute to the value of SMIS to the intended end users.

2. Literature review methodology

2.1. Scope of the review

On behalf of the HDC, Rickson and Deeks (2013) conducted a review of soil management challenges and practices in horticulture (CP107a). The purpose of that review was to identify key gaps in research and in knowledge transfer mechanisms that hinder the development and implementation of best practice guidelines for sustainable soil management. As a result of the grower survey undertaken as part of this gap analysis, 11 specific soil management challenges were identified (

Table 11).

Table 11. Soil management challenges in horticulture, identified in CP107a (Rickson and Deeks, 2013)

Soil Management Challenges in Horticulture		
Soil erosion by water	Drought	Yield quantity
Soil compaction	Accessing wet soils	Yield reliability
Soil erosion by wind	Drainage	Yield quality
Too little organic matter	Soil-borne diseases	

To reflect end user requirements of SMIS, these identified issues formed the framework / structure of CP107d's literature review. These key soil management challenges were supplemented by other soil management factors considered to be important for inclusion in SMIS by the project team and Steering Committee members. These additional soil management factors included nutrient management, pest control, soil biodiversity, weeds and soil acidity/alkalinity.

2.2. Sources of information

Sources of information covered in the review included:

- Academic papers published in scientific, peer reviewed journals;
- Conference proceedings / papers;
- Research reports;
- Grey literature (e.g. articles on websites and in trade magazines)

Collectively, these sources are termed 'literature' in the following sections. Each source (or 'item') was classified as being 'quantitative', 'qualitative' and 'anecdotal' (see accompanying Excel spreadsheet). This classification was used to evaluate and quantify the confidence in outputs / findings from each item (i.e. the 'weight of evidence' within the SMIS database). It is envisaged that a weighting system needs developing to reflect the confidence behind all the different data sources within SMIS (i.e. grower data [GK, Muddy Boots], literature, research projects, expert opinion: See Annual Report, November 2016 for further detail of information / data / knowledge sources to be used in SMIS). The Cranfield team will consider how an evidence weighting system can be incorporated into the SMIS system architecture (e.g. through visualisation). The milestone for achieving this is 30/11/17 (see meeting minutes of 31/03/17).

2.3. Analysis of sources

The review of the literature had two main phases (Figure 12). The first phase required the collation of literature relevant to horticultural soil management. This brought together literature from two sources:

the literature identified in the gap-analysis review conducted by Rickson and Deeks (2013).

a new literature search (detailed below) that updated Rickson and Deeks' (2013) literature with relevant published papers between 2013 and 2016.

The second phase was then to review the collated literature from both sources in order to identify, categorise and catalogue case studies that had potential in providing knowledge that can be incorporated into the SMIS database (Figure 12).

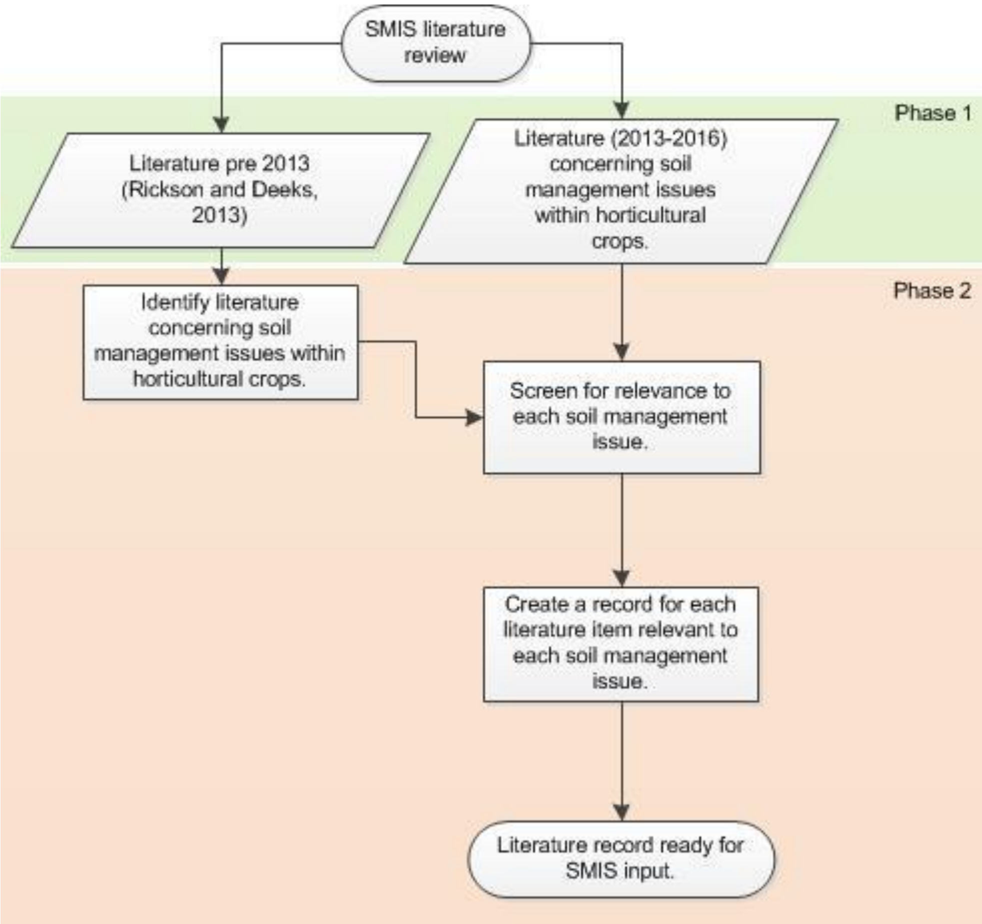


Figure 12. Literature review process

2.4. Phase 1

All 375 items of literature referenced by Rickson and Deeks (2013) were collated and uploaded to Mendeley (a reference management software; https://www.mendeley.com/; Figure 2). In order to update this literature, a search for additional literature published since 2013 (i.e. 2013 to 2016) was also undertaken.

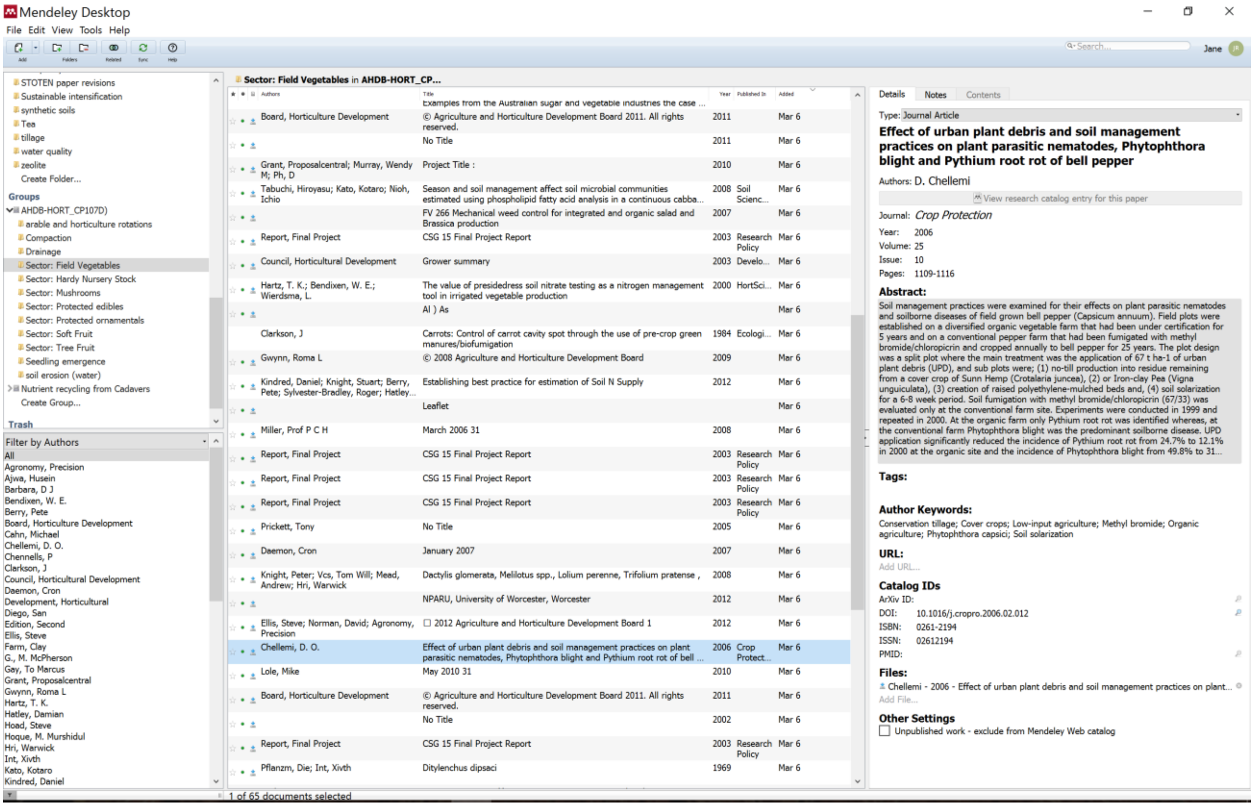


Figure 13. Screen shot of Mendeley reference management software. Extract shown is the list of literature related to soil management in Field Vegetables. Panel on the right shows details of an individual literature item, including abstract, data generated, key words and full citation.

For academic literature (including papers and conference proceedings), a search was conducted using Scopus (http://www.scopus.com/home.url). The search terms trialed for this search are presented in Table 12. The aim was to select a search term that was not too broad (exhaustive and time consuming) and not too restrictive (with the risk that not all relevant articles would be captured). The keyword search results were considered to be most representative of relevant literature (Table 12), giving a total of 206 items of academic literature. The citations and abstracts were exported into Mendeley, a reference management software.

Table 12. Search terms trialled in Scopus (August 2016) and the number of hits. *denotes wild cards.

Search term(s)	In abstract	In title	In keywords	Comment
(("Soil") AND ("horticultur*"))	639	21	52	Search term too broad in abstract and titles, but will capture all relevant papers
(("Soil") AND ("manage*") AND ("horticultur*"))	174	1	13	Search term too broad, but with no papers with search term in the title
(("Soil manage*") AND ("horticultur*"))	16	0	1	Search term reasonable, but with no papers with search term in the title
(("Soil") AND ("manage*") AND ("vegetable*"))	473	9	68	Search term too broad in abstract, with too few papers with relevant title
(("Soil") AND ("manage*") AND ("fruit*"))	676	5	71	Search term too broad, with too few papers with relevant title
(("Soil") AND ("manage*") AND ("mushroom*"))	25	0	4	Search term too narrow in abstract, with no papers with relevant title
(("Soil") AND ("manage*") AND ("protected crop*"))	5	0	0	No papers with search term in the abstract or title
Totals	2008	36	206	

For literature arising from research projects (e.g. final reports), UK Research Council and funding bodies (BBSRC, NERC, AHDB, Defra etc.) research outputs were accessed online. Also, soil management projects started since 2013 were sought and screened for relevance. Relevant projects within AHDB Horticulture but presently without full access to final reports and their Appendices were noted for later follow-up. Research projects included by Rickson and Deeks (2013), that were not fully completed by 2013, were updated with review of final reports where available. Research projects were also uploaded to Mendeley.

Grey literature searches were undertaken outside of Mendeley using Google's search engine. Here the search string used included 'horticulture' and 'soil' and 'management', followed by each soil management issue in turn (e.g. compaction, loss of organic matter etc.). The first 5 pages of each item were reviewed for relevance to SMIS and end user requirements.

2.5. Phase 2

Using Mendeley's analytical functionalities, an in-text search was undertaken on all literature items for each of the identified soil management challenges, using pre-selected search terms (**Table 13**). These were deemed appropriate search terms for the identification of literature on best management practices that mitigate against the identified soil management challenges and associated crop yield effects. This considerably reduced the number of literature items down to approximately 207.

Table 13. Soil management issues and their respective search terms.

Soil management challenge	Search term(s)
Soil erosion by water/wind	Erosion
Soil compaction	Compact*
Too little organic matter	Organic matter / carbon
Soil water management (incorporating drought, drainage and accessing wet soils)	Water / moisture
Soil-borne disease	Disease*
Pests	Pest*
Nutrient management	Nutrient*
Soil biodiversity	Biodivers* / diversity
Weeds	Weed*
Soil acidity/alkalinity	pH

*denotes multiple endings are possible.

Literature ‘hits’ were separated by each soil management challenge, and closely scrutinised to identify whether the search term ‘hit’ also referred to best soil management practices that addressed that challenge (i.e. soil management solution). Focus was also given to literature pertaining to temperate climates, and agricultural practices similar to that undertaken in the UK horticultural rotations (e.g. studies on soil management issues and practices in rice paddy systems was excluded). Where a full-text of the literature was not available, inter-library loans were made through Cranfield University’s Information Services.

Each item of literature was classified by knowledge type; quantitative (based on empirical evidence from field work: laboratory studies were excluded due to the limitations of extrapolating practical, applied results from small spatial scales); qualitative (based on observations during a field-based experiment); and anecdotal (unreferenced statements).

It was assumed that items with quantified data would provide more confidence to end users than qualitative or anecdotal information (See earlier comment about developing a ‘weighting system’ in the SMIS architecture). Therefore, for each soil management issue, the specific details of available quantitative knowledge within each item of literature were extracted into a common descriptive form (termed meta-criteria in the Annual Report 2016) (**Table 14**). The purpose was to first standardise the data extracted from the diverse range of sources into a common format. This will aid input of that data into the SMIS architecture. Examples of items with quantified data are available in the accompanying Excel file. Extracted data included (where available) both details of the degree of the soil management challenge, the impact of soil management practices on the soil management issue and the subsequent impact on yield (quality, quantity and reliability). In addition, comment is made of the value added to SMIS from the data contained in each item, and how this will benefit the intended end users of SMIS.

Table 14. Details of the fields contained in the meta-criteria table.

Heading	Criteria contained	Justification for inclusion
Document reference	Literature citation	
AHDB Horticulture Sector	e.g. Field vegetables, bulbs, soft fruits, brassicas	Key for categorising knowledge.

Soil management challenge	e.g. erosion, compaction, loss of organic matter	Key soil management issues / challenges identified by growers (see CP107a)
Soil management solution(s)	e.g. ground cover, reduced tillage, mulching	Identifies where soil management practice has been tested (and outcomes)
Place	Geolocation: field grid reference and geometry; farm map Field size and partitioning: Organic/conventional Grower address; Cadastral; Parcel history; Rented/Owned	Evidences the geographical spread of the literature and of the soil management issue.
Experimental setup	No. sites Plot size No. replicates	Evidences how the robust the knowledge is, can help in later weighting.
Time	Date Interval Season	Evidences the temporal spread.
Characteristics	Soil properties Climate and weather patterns Position in landscape/topography	Evidences the geographical spread of the literature and soil management issues.
Land Use	Crop type and variety; companion cropping Planting date; seeding rate; harvest date Rotational context	Key for categorising the available knowledge.
Operations	Soil management field operations Organic amendments Use of field buffer strips	Evidences the treatments tested and practices in place.
Outcomes	Yield (quality, quantity and reliability) and productivity; soil nutrient status; soil degradation control and remediation.	The key message to the SMIS user

	Harvest date; pests and diseases Water management; drainage status	
Datasets	Y / N	Helps identify and categorise available data sources.
Reservations	Notes relevant to study factors.	Pulls out any concern regarding the integrity of the research e.g. lack of replication.

3. Literature results

In total, 581 items of literature were reviewed (full list is available). In total, 84 items were considered relevant to SMIS. This has dropped since the last update (Meeting of the 31st March 2017, Cranfield University) as laboratory studies and results from non-temperate climates have been excluded.

A total of 84 relevant items of literature were identified through this review. Within these, 172 knowledge items for horticultural soil management have been found. (This larger number is because some papers deal with more than one soil management issue and/or practice. For example, the paper by Bonanomi et al. (2008) deals with control of weeds in protected edibles by a number of soil management practices: the use of soil solarisation, plastic mulches and soil amendments.) The greatest number of literature items focus on research undertaken in the UK, with a good global distribution of other literature (Figure 14). Field vegetable, tree fruit, cross sector and protected edibles are particularly well represented (Figure 15). Quantitative references to soil management challenges are the most frequent, making up 61 % of the identified knowledge sources (Figure 16). The greatest amount of literature was found for soil-borne disease, followed by weeds and nutrient supply (Figure 17). The literature covers 20 broad soil management solutions with a focus towards generic ‘management practice’ and rotation based solutions (Figure 18). All other solutions are evenly distributed across the identified literature.

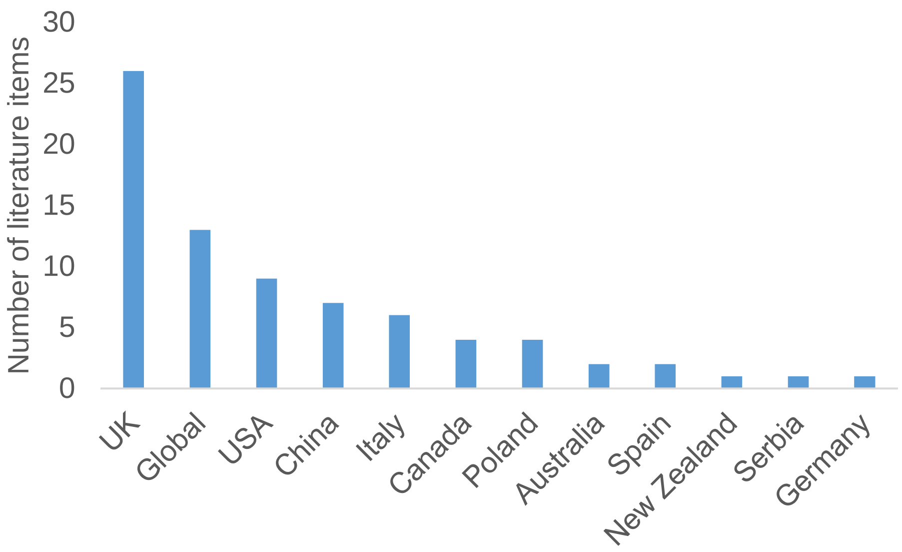


Figure 14. Distribution of literature identified by country.

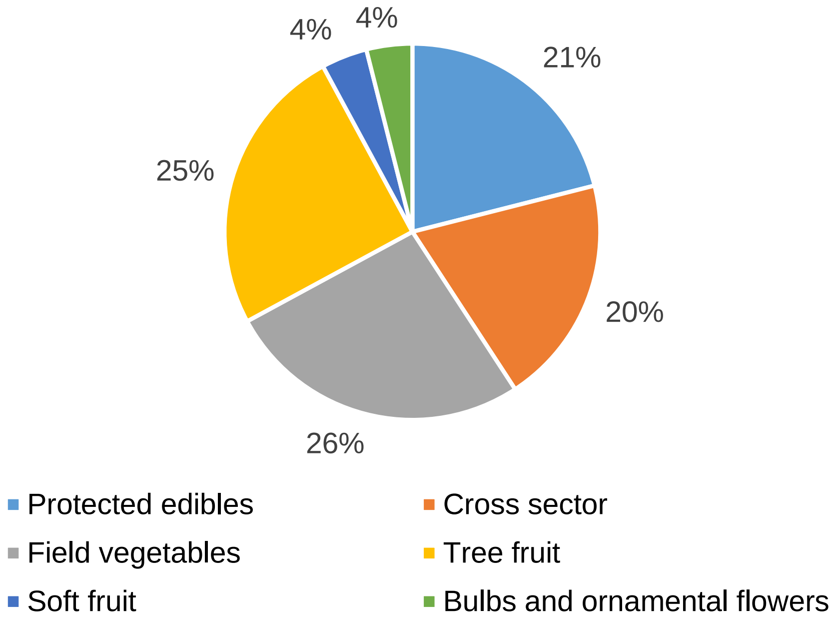


Figure 15. The distribution of relevant literature by AHDB Horticulture sector.

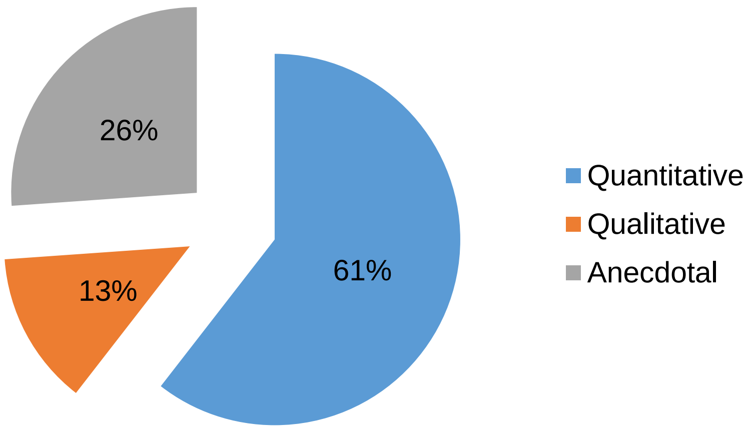


Figure 16. Classification of knowledge sources within the identified literature.

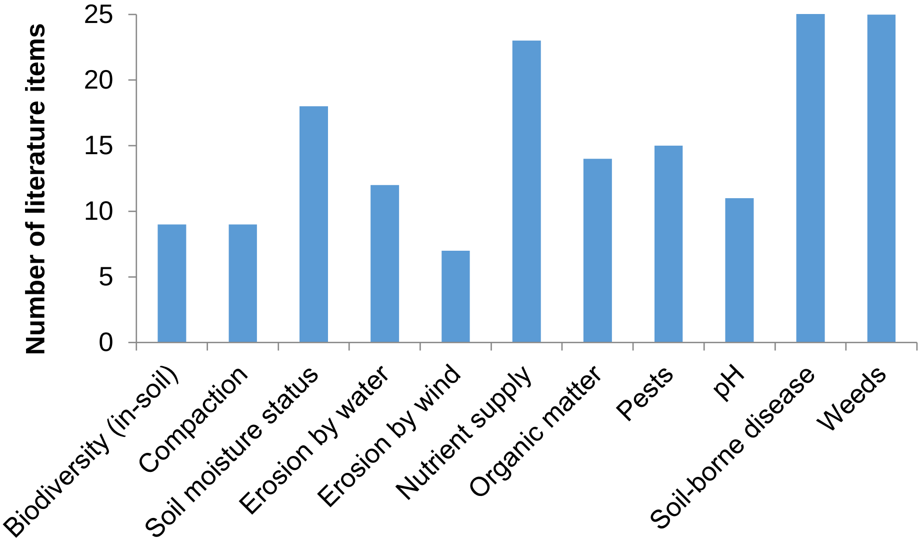


Figure 17. Distribution of literature identified relative to soil management challenge.

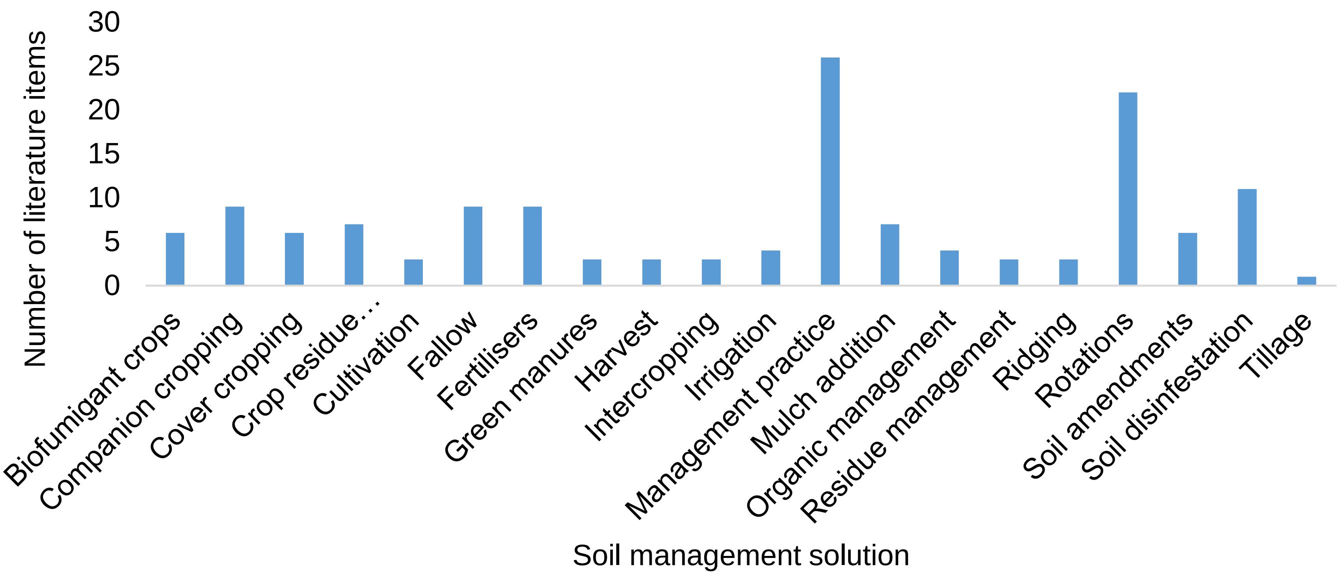


Figure 18. Distribution of literature identified relative to soil management solution(s).

4. Conclusion

The literature identified in this review forms a sound and integrated basis of horticultural soil management knowledge. Benefits to the SMIS end users include:

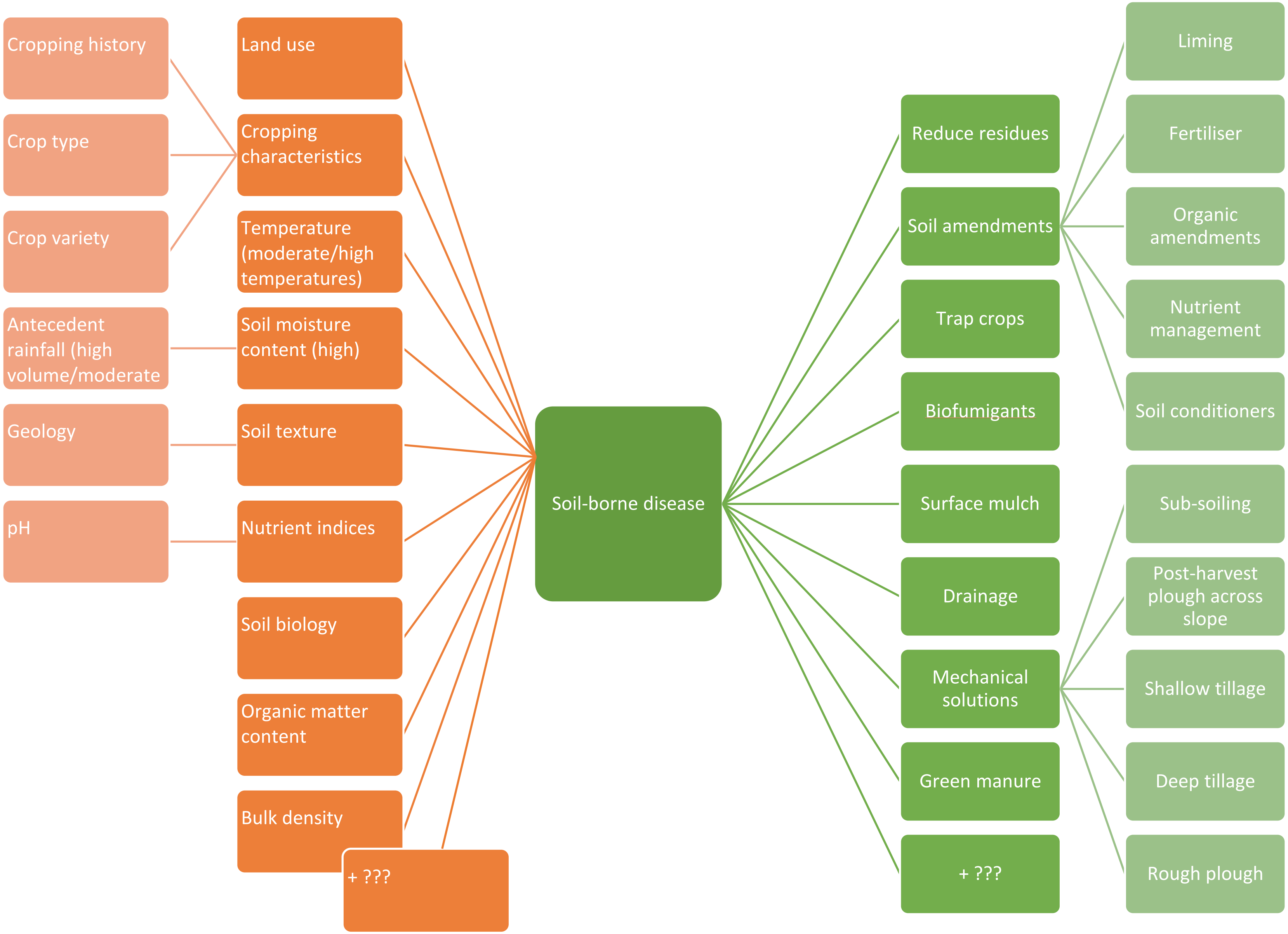
- a unique, novel and up-to-date synthesis of the extensive and diverse research outputs related to horticultural soil management issues and solutions
- easy access to knowledge that has previously been inaccessible (e.g. locked up in subscription-based academic papers)
- expansion of the SMIS database in terms of quantitative, qualitative and anecdotal knowledge on both horticultural soil management issues and management practices used to both prevent and remediate these issues.

Having extracted the relevant information from the literature review in a systematic way (e.g. see Example tabs in the accompanying Excel spreadsheet), the next step is to work with Cranfield colleagues on the informatics team to understand how the knowledge identified in the literature review (in quantitative, qualitative and / or anecdotal form) can be accommodated in the SMIS database, alongside the grower data (e.g. Gatekeeper records) and expert knowledge/ opinions (e.g. outputs from the Fuzzy Cognitive Mapping exercise). The vision is for the resulting SMIS database to be interrogated to identify best soil management practices to:

- alleviate and prevent soil degradation in horticultural crop production

- reduce input costs associated with remediation of soil management problems in horticulture
- remove the soil-based constraints that currently limit optimal crop production in horticulture; and
- enhance soil quality and soil health to bring agronomic and financial benefits to the grower.

Appendix 3. Soil management causes and solutions: Soil borne diseases



Appendix 4. Relevant literature colour coded and categorised.

Table 4.1. Literature identified as relevant to SMIS including the soil management challenge and management solution addressed. Each of these are categorised by knowledge type: Quantitative (green); Qualitative (yellow); Anecdotal (blue).

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Abdul-Baki, A. A. Teasdale, J. R. Goth, R. W. and Haynes, K. G. (2002) Marketable yields of fresh-market tomatoes grown in plastic and hairy vetch mulches	USA	Tomato		x	x	x	x	x					x			x																	x

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Alexander, P. D. and Nevison, I. M. (2015) 'The long-term effects of repeated application of the same organic material to soil in a horticultural context', *Acta Horticulturae*, 1076, pp. 143–150.	UK	Ornamental plants						x	x		x																				x		
Arancon, N. Q. Edwards, C. A. and Lee, S. S. (1998) Management of plant parasitic nematode population by use of vermicomposts	USA	Tomato, bell pepper,			x							x																			x		

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Bailey, K. . and Lazarovits, G. (2003) 'Suppressing soil-borne diseases with residue management and organic amendments', *Soil and Tillage Research*, 72(2), pp. 169–180.	Global	Cross sector						x				x					x	x													x		
Barbara, J. (2010) FV 353 - Carrot cavity spot: (i) using quantitative PCR to predict disease in strawed crops; (ii) controlling moisture for optimum disease management	UK	Carrot										x												x									

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Beniston, J. W., Lal, R. and Mercer, K. L. (2016) 'Assessing and Managing Soil Quality for Urban Agriculture in a Degraded Vacant Lot Soil', *Land Degradation & Development*. John Wiley and Sons Ltd, 27(4), pp. 996–1006.	USA	Tomato, swiss chard		x					x							x															x		

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Beslic, Z., Pantelic, M., Dabic, D., Todic, S., Natic, M. and Tesic, Z. (2015) 'Effect of vineyard floor management on water regime, growth response, yield and fruit quality in Cabernet Sauvignon', *Scientia Horticulturae*. Elsevier, 197, pp. 650–656.	Serbia	Vineyards			x										x																		x

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Biala, J. and Milgate, M. (2014) 'Grower expectations and experiences with the use of organic mulches and soil amendments in the horticultural industry in Queensland, Australia', *Acta Horticulturae*, 1018, pp. 473–480.	Australia	Perennial fruit/veg						x	x				x													x					x		
Biddlecombe, T. (2010) TF 179 — Pear: The effect of soil moisture on fruit storage quality	UK	Pear			x																					x							

References / Bibliography	Country	Crop	Soil management challenge (see CP107a): Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds
Biddlecombe, T. (2012) TF 177 - Apples: Long term effects of applied composted green waste mulch on the cropping of Braeburn and Cox	UK	Apple			x			x	x	x	x		x
Bonanomi G., Antignani V., Capodilupo M. and Scala F. (2010). Identifying the characteristics of organic soil amendments that suppress soilborne plant diseases. Soil Biology and Biochemistry 42, 136-144.	Global	Cross sector										x	

References / Bibliography	Country	Crop	Soil management solutions: Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Biddlecombe, T. (2012) TF 177 - Apples: Long term effects of applied composted green waste mulch on the cropping of Braeburn and Cox	UK	Apple													x							
Bonanomi G., Antignani V., Capodilupo M. and Scala F. (2010). Identifying the characteristics of organic soil amendments that suppress soilborne plant diseases. Soil Biology and Biochemistry 42, 136-144.	Global	Cross sector				x														x		

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Bonanomi G., Chiurazzi M., Caporaso S., Del Sorbo G., Moschetti G. and Felice S. (2008). Soil solarization with biodegradable materials and its impact on soil microbial communities. Soil Biology and Biochemistry 40, 1989-1998.	Italy	Tomato	x							x		x	x													x					x	x	
Bond, W. and Grundy, A. C. (2001) Non-chemical weed management in organic farming systems	Global	Cross sector											x			x							x			x				x			x

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Brainard, D. C., Bakker, J., Noyes, D. C. and Myers, N. (2012), "Rye Living Mulch Effects on Soil Moisture and Weeds in Asparagus", HortScience, vol. 47, no. 1, pp. 58-63.	USA	Asparagus		x	x	x		x		x			x		x																		

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Campiglia, E., Radicetti, E. and Mancinelli, R. (2015) 'Cover crops and mulches influence weed management and weed flora composition in strip-tilled tomato (Solanum lycopersicum)', *Weed Research*. Edited by C. Kempenaar. Blackwell Publishing Ltd, 55(4), pp. 416–425.	Italy	Tomato											x			x																	

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Castro, E., Mañas, P. and De Las Heras, J. (2013) 'Effects of wastewater irrigation in soil properties and horticultural crop (*LACTUCA SATIVA* L.)', *Journal of Plant Nutrition*, 36(11), pp. 1659–1677.	Spain	Lettuce						x	x															x									

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)										
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds
Chellemi, D. 0. (2006) Effect of urban plant debris and soil management practices on plant parasitic nematodes, Phytophthora blight and Pythium root rot of bell pepper. *Crop Protection.* Vol. 25, no. 10, pp. 1109-1116	USA	Bell pepper								x		x	
Clarkson, J. (2012) FV 405 — Carrots: Control of carrot cavity spot through the use of pre-crop green manures/biofumigation	UK	Carrot										x	

References / Bibliography	Soil management solutions																			
	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Chellemi, D. 0. (2006) Effect of urban plant debris and soil management practices on plant parasitic nematodes, Phytophthora blight and Pythium root rot of bell pepper. *Crop Protection.* Vol. 25, no. 10, pp. 1109-1116				x									x							
Clarkson, J. (2012) FV 405 — Carrots: Control of carrot cavity spot through the use of pre-crop green manures/biofumigation								x											x	

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Crecchio, C. Gelsomino, A. Ambrosoli, R. Minati, J. L. and Ruggiero, P. (2004) Functional and molecular responses of soil microbial communities under differing soil management practices	Italy	Non-specific	X																X						X		X						X
Davies, J. (2007) PC 263 - Protected lettuce: to assess the effects of soil nitrogen levels, trickle irrigation and light levels on nitrate accumulation in winter crops.	UK	Lettuce						X																X									

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Defra. (2005) Controlling soil erosion	UK	Cross sector				x	x								x	x										x							x
Defra. (2009) Protecting our water, soil and air — A code of good agricultural practice for farmers, growers and land managers	UK	Cross sector						x	x							x	x	x		x						x					x		

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Dorais, M. and Alsanius, B. (2015) 'Advances and trends in organic fruit and vegetable farming research', in Janick, J. (ed.) *Horticultural Reviews*. Hoboken, New Jersey: John Wiley & Sons, Inc., pp. 185–268.	Global	Vegetables	x										x														x						

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Döring, J., Frisch, M., Tittmann, S., Stoll, M. and Kauer, R. (2015) 'Growth, Yield and Fruit Quality of Grapevines under Organic and Biodynamic Management.', *PloS one*. Public Library of Science, 10(10), p. e0138445.	Germany	Grape	x			x	x	x	x			x															x						
Edwards-Jones, G. (2010) CP 62 - Carbon storage in orchards and amenity plantings	UK	Orchards			x	x																			x								

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Ferris, H. Venette, R. C. and Scow, K. M. (2004) Soil management to enhance bacterivore and fungivore nematode populations and their nitrogen mineralisation functions	USA	Tomatoes								x						x																	

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Forge, T. A. Hogue, E. Nielsen, G. and Nielsen, D. (2003) Effects of organic mulches on soil microfauna in the root zone of apple: implications for nutrient fluxes and functional diversity of the soil food web (also available as a paper Forge et al 2003)	Canada	Apple						x																		x							
Gardner, C. M. K., Laryea, K. B. and Unger, P. W. (1999) Soil Physical Constraints To Plant Growth and Crop Production.	Global	Cross sector		x		x	x						x				x									x							

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Ge, T., Chen, X., Yuan, H., Li, B., Zhu, H., Peng, P., Li, K., Jones, D. L. and Wu, J. (2013) 'Microbial biomass, activity, and community structure in horticultural soils under conventional and organic management strategies', *European Journal of Soil Biology*. Elsevier Masson SAS, 58, pp. 122–128.	China	Cabbage, tomato	x					x	x																x		x						
Gladders, P. (1997) FV 167a — Parsnip: control of canker	UK	Parsnip								x		x	x					x															

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Gladders, P. Davies, G. Wolfe, M and Haward, R. (2002) Development of disease control strategies for organically grown field vegetables (DOVE).	UK	Vegetable crops										X	X													X					X		
Goh, K. M. Pearson, D. R. and Daly, M. J. (2001) Effects of apple orchard production systems on some important soil physical, chemical and biological quality parameters	New Zealand	Apple		X				X			X															X		X					

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Gonzalez Perez, P., Ye, J., Wang, S., Wang, X. and Huang, D. (2014) 'Analysis of the occurrence and activity of diazotrophic communities in organic and conventional horticultural soils', *Applied Soil Ecology*. Elsevier B.V., 79, pp. 37–48.	China	Spinach, pepper, onion, celery,	x							x		x											x		x								

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Grassi, F., Mastrorilli, M., Mininni, C., Parente, A., Santino, A., Scarcella, M. and Santamaria, P. (2014) 'Posidonia residues can be used as organic mulch and soil amendment for lettuce and tomato production', *Agronomy for Sustainable Development*. Springer-Verlag France, 35(2), pp. 679–689.	Italy	Lettuce, tomato			x	x	x	x	x																	x							
Grundy, A. (2007) FV 266 — Mechanical weed control for integrated and organic salad brassica production	UK	Lettuce											x																				x

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Guzzon, R., Gugole, S., Zanzotti, R., Malacarne, M., Larcher, R., von Wallbrunn, C. and Mescalchin, E. (2016) 'Evaluation of the oenological suitability of grapes grown using biodynamic agriculture: the case of a bad vintage.', *Journal of applied microbiology*. Blackwell Publishing Ltd, 120(2), pp. 355–65.	Italy	Grapes										x													x								

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Hanks, G. et al., (2016). BOF 076a - Understanding the physiological disorders in daffodil - BOF 076 project extension to study the three-year-down crop.	UK	Daffodil			X							X								X										X			
Haynes, R. J. (1980) 'Influence of soil management practice on the orchard agro-ecosystem', *Agro-Ecosystems*, 6(1), pp. 3–32.	Global	Fruit			X			X		X		X				X		X		X													

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Herold, L. et al. (2017) FV 428 - Vining peas: The effect of soil phosphate levels on rhizobial population.	UK	Peas	x																	x													
Hoque, M. M. Ajwa, H. Othman, M. Smith, R. and Cahn, M. (2010) Yield and postharvest quality of lettuce in response to nitrogen, phosphorus and potassium fertilizers. HortScience, 45 (10): 1539-1544	Canada	Lettuce						x												x													
Horticulture New Zealand (2015). Soils/erosion environment guide.	New	Squash,					x								x																		

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)												Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds		Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Kawecki, Z. and Tomaszewska, Z. (2006) The effect of various soil management techniques on growth and yield in the black chokeberry (Aronia melanocarpa Elliot)	Poland	Black chokeberry											x																			x		

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Keesstra, S., Pereira, P., Novara, A., Brevik, E. C., Azorin-Molina, C., Parras-Alcántara, L., Jordán, A. and Cerdà, A. (2016) 'Effects of soil management techniques on soil water erosion in apricot orchards.', *The Science of the total environment*. Elsevier, 551–552, pp. 357–66.	Spain	Apricot			x	x			x							x																	x

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Knight, P. (2005) FV 249 — Carrots: A cost-benefit study in the control of free-living nematodes, cavity spot and volunteer potatoes by comparing specific management systems before and during cropping	UK	Carrot								x		x	x						x		x												
Knight, P. (2008) FV 299 - An investigation into the adoption of green manures in both organic and conventional rotations to aid nitrogen management and maintain soil structure.	UK	Bulb onion, barley, potatoes			x			x	x	x			x								x												

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																				
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Knight, P. (2009) FV 299a - Extension of FV 299 — investigation into the adoption of green manures in both organic and conventional rotations to aid nitrogen management and maintain soil structure	UK	Bulb onion		x				x	x	x			x								x												
Knott, C. (2011) FV 372 - Evaluation of potential alternatives for weed control in asparagus following the loss of herbicides.	UK	Asparagus											x									x											

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Koehn, A. C. Peryea, F. J. Neilsen, D. and Hogue, E. J. (2002) Temporal changes in nitrate status of orchard soils with varying management practices	USA	Apple						x																		x							
Lazzeri, L. Baruzzi, G. Malaguti, L. and Antoniacci, L. (2003) Replacing methyl bromide in annual strawberry production with glucosinolate-containing green manure crops LEAF. (2011) Simply sustainable soils	Italy	Strawberry								x		x	x	x							x												

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Leandro, L. F. S. Guzman, T. Ferfuson, L. M. Fernandez, G. E. and Louws, F. J. (2005) Population dynamics of fungi associated with strawberry roots in different soil management systems (also available as paper 2007)	USA	Strawberry										x																			x		

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Li, H. C., Gao, X. D., Zhao, X. N., Wu, P. T., Li, L. S., Ling, Q. and Sun, W. H. (2016) 'Integrating a mini catchment with mulching for soil water management in a sloping jujube orchard on the semiarid Loess Plateau of China', *Solid Earth*. Copernicus GmbH, 7(1), pp. 167–175.	China	Jujube			x																					x							

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Liu, L., Sun, C., Liu, S., Chai, R., Huang, W., Liu, X., Tang, C. and Zhang, Y. (2015) 'Bioorganic Fertilizer Enhances Soil Suppressive Capacity against Bacterial Wilt of Tomato', *PLOS ONE*. Edited by G. Zhang. Public Library of Science, 10(4), p. e0121304.	China	Tomato	x									x								x													

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Machado, A., Jakelaitis, A., Ferreira, L., Agnes, E. and Santos, L. (2005) 'Population Dynamics of Weeds in No-Tillage and Conventional Crop Systems', Journal of Environmental Science and Health, Part B: Pesticides, Food Contaminants, and Agricultural Wastes, 40(1), pp. 119–128.	Brazil	Maize, beans											x																				x
Magdoff, F. and Van Es, H. (2009) Building soils for better crops — sustainable soil management	Global	Cross sector				x	x		x																								

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Neilsen, G. H. Hogue, E. J. Forge, T. and Neilsen, D. (2003)a Mulches and biosolids affect vigour, yield and leaf nutrition of fertigated high density apple	Canada	Apple			x						x		x													x					x		
Neilsen, G. H. Hogue, E. J. Forge, T. and Neilsen, D. (2003)b Surface application of mulches and biosolids affect orchard soil properties after 7 years	Canada	Apple			x			x	x		x		x													x					x		

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Neilson, R. (2016). FV 447 - Carrots & Parsnips - Developing a strategy to control Free Living Nematodes.	Global	Carrots, parsnip								x				x							x										x		x
Niziolomski, J.C. 2011. Critical evaluation of the role of sub-soiling and mulching in the control of runoff and erosion from asparagus furrows. (Unpublished Master of Science Thesis) Cranfield University, UK.	UK	Asparagus		x		x																				x							x

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Noble, R. (2011) FV 359 - Nutrient Requirements for Field Grown Herbs.	UK	Coriander, mint						x												x													
O'Neil, T. (2009) 'A Summary Report of Useful Information on Soil Disinfestation Gathered for UK Growers'.	Global	Cross sector											x																			x	
O'Neil, T. (2010) HNS 137 - Field grown trees: evaluation of chemical and biological pre-plant soil treatments for control of Verticillium wilt.	UK	Maple, lime, whitebeam										x									x												

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Powlson, D. S., Gregory, P. J., Whalley, W. R., Quinton, J. N., Hopkins, D. W., Whitmore, A. P., Hirsch, P. R. and Goulding, K. W. T. (2011) 'Soil management in relation to sustainable agriculture and ecosystem services', *Food Policy*, 36, pp. S72–S87.	Global	Cross sector			x													x															

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Rickson, J., Deeks, L., Posthumus, H. and Quinton, J. (2010) To review the overall costs and benefits of soil erosion measures and to identify cost-effective mitigation measures. Sub-Project C of Defra Project SP1601: Soil Functions, Quality and Degradation — Studies in Support of the Implementation of Soil Policy.	UK	Cross sector				x	x																										
Runham, S. R. (1993) FV 080 — Bulb onions: control of wind erosion on light soils	UK	Onion					x								x																		

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Ruzzi, M. and Aroca, R. (2015) 'Plant growth-promoting rhizobacteria act as biostimulants in horticulture', *Scientia Horticulturae*. Elsevier, 196, pp. 124–134.	Global	Cross sector			x			x																							x		
SoilCare (2007). Northern Rivers soil BMP Guide — Perennial Horticulture — NZ	New	Generic	x	x		x									x										x								

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Stirling, R. (2008) The impact of farming systems on soil biology and soilbome diseases: examples from the Australian sugar and vegetable industries — the case for better integration of sugarcane and vegetable production and implications for future research	Australia	Sugar cane and vegetables		x	x	x		x		x	x	x	x															x		x	x		x
Szewczuk, A. and Gudarowska, E. (2004) The effect of soil mulching and irrigation on yielding of apple trees in ridge planting	Poland	Apple											x													x							

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Szewczuk, A. and Gudarowska, E. (2006) Performance of young peach trees planted in ridges and mulched in tree rows	Poland	Peach			x								x													x							
Tompsett, A. (2009) BOF 050 — Narcissus: overcoming the problem of soil sickness with particular reference to the Isles of Scilly	UK	Narcissus								x		x		x																			

References / Bibliography	Country	Crop	Soil management challenge (see CP107a): Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Soil management solutions: Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Treder, W. Klamkowski, K. Mika, A. and Wojcik. (2004) Response of young apple trees to different orchard floor management systems	Poland	Apples			x								x													x							
Wallace, P. and Carter, C. (2007) Effects of compost on yields of winter wheat and barley, sugar beet, onion and swede in the fourth and fifth years of a rotation	UK	Sugarbeet			x	x		x	x		x															x							

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Wang, H., Wang, C., Zhao, X. and Wang, F. (2015) 'Mulching increases water-use efficiency of peach production on the rainfed semiarid Loess Plateau of China', *Agricultural Water Management*. Elsevier, 154, pp. 20–28.	China	Peach			x																					x							

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Wang, X.-L., Ye, J., Perez, P. G., tang, D.-M. and Huang, D.-F. (2013) 'The impact of organic farming on the soluble organic nitrogen pool in horticultural soil under open field and greenhouse conditions: a case study', *Soil Science and Plant Nutrition*, 59(2), pp. 237–248.	China	Field veg including lettuce.						x			x														x		x						
Washington State University Extension (Accessed 2016) Production: Soil management for in-ground nursery crops.	USA	Nursery stock							x																			x					

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Wedgwood, E. and Hamilton, M. (2016). FV 449 - Onions: Investigation into the control of White Rot in bulb and salad onion crops	UK	Onion										x									x										x	x	
Wilson, M. (2009) CP 006 — Integrated use of soil disinfection and microbial organic amendments for the control of soil borne diseases and weeds in sustainable crop production (HortLINK)	UK	Non-specific	x																												x		

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)										
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds
Yang, L., Zhao, F., Chang, Q., Li, T. and Li, F. (2015) 'Effects of vermicomposts on tomato yield and quality and soil fertility in greenhouse under different soil water regimes', *Agricultural Water Management*. Elsevier, 160, pp. 98–105.	China	Tomato						x	x				

References / Bibliography	Soil management solutions																			
	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Yang, L., Zhao, F., Chang, Q., Li, T. and Li, F. (2015)													x							

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Yao, S. Merwin, I. A. Bird, G. W. Abawi, G. S. and Thies, J. E. (2005) Orchard floor management practices that maintain vegetative or biomass groundcover stimulate soil microbial activity and alter soil microbial community composition	USA	Apple	x					x			x															x							
Young (2012) FV 361 — Reducing the impact of sclerotinia disease on arable rotations, vegetable crops and land use	UK	Peas, beans.										x																					x

References / Bibliography	Country	Crop	Soil management challenge (see CP107a)											Soil management solutions																			
			Biodiversity (in-soil)	Compaction	Soil moisture status	Erosion by water	Erosion by wind	Nutrient supply	Organic matter	Pests	pH	Soil-borne disease	Weeds	Biofumigant crops	Companion cropping	Cover cropping	Crop residue	Cultivation	Fallow	Fertilisers	Green manures	Harvest	Intercropping	Irrigation	Management practice	Mulch addition	Organic management	Residue management	Ridging	Rotations	Soil amendments	Soil disinfestation	Tillage
Zaidi, A., Ahmad, E., Khan, M. S., Saif, S. and Rizvi, A. (2015) 'Role of plant growth promoting rhizobacteria in sustainable production of vegetables: Current perspective', *Scientia Horticulturae*. Elsevier, 193, pp. 231–239.	Global	Cross sector						x																							x		

Appendix 5. 2015 rotational context of vining pea growers

Grower AY-LM
1097(ha)

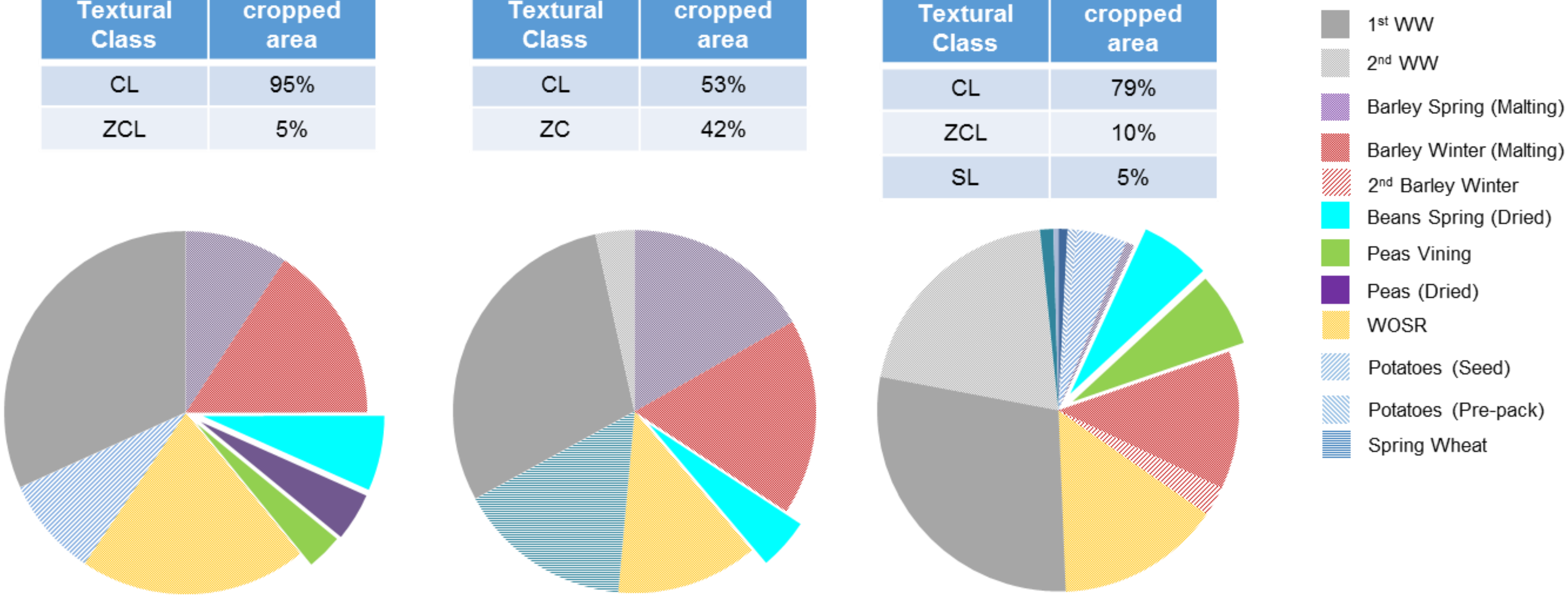
Soil Textural Class	% of cropped area
CL	95%
ZCL	5%

Grower AY-R
967 (ha)

Soil Textural Class	% of cropped area
CL	53%
ZC	42%

Grower JSR
3563 (ha)

Soil Textural Class	% of cropped area
CL	79%
ZCL	10%
SL	5%



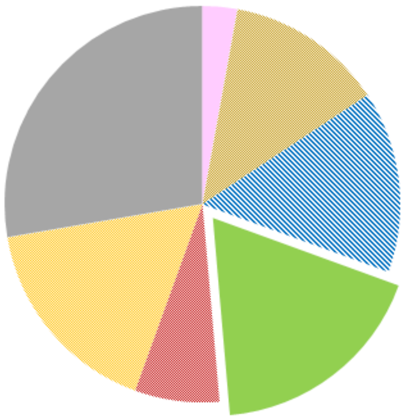
Horticultural Crops (155 ha)

Horticultural Crops (44.2 ha)

Horticultural Crops (450 ha)

Grower AN
970 (ha)

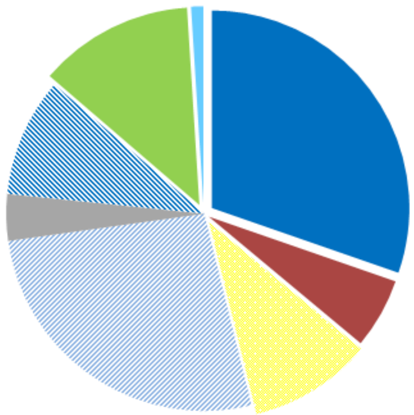
Soil Textural Class	% of cropped area
SL	94%
CL	6%



Horticultural Crops
(175 ha)

Grower JBF
327 (ha)

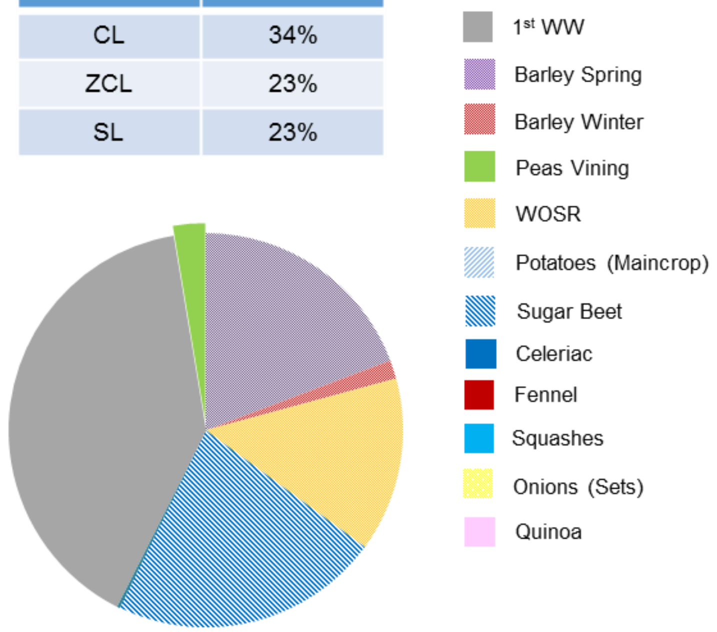
Soil Textural Class	% of cropped area
ZCL	99%
ZC	1%



Horticultural Crops
(196 ha)

Grower CF
2607 (ha)

Soil Textural Class	% of cropped area
CL	34%
ZCL	23%
SL	23%



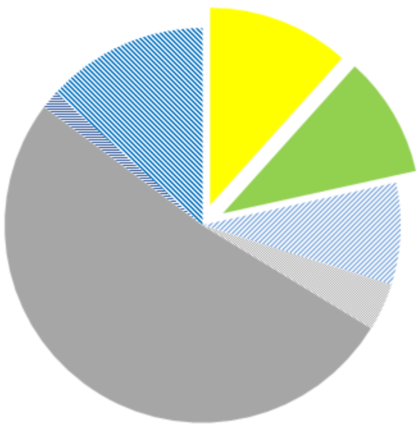
Horticultural Crops
(68 ha)

Grower HF
1215 (ha)

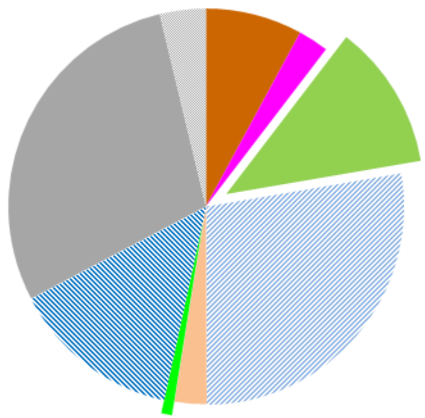
Soil Textural Class	% of cropped area
ZCL	72%
ZC	28%

Grower WF
2144 (ha)

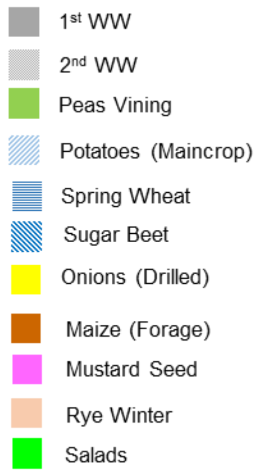
Soil Textural Class	% of cropped area
ZCL	95%
C	5%



Horticultural Crops
(261 ha)



Horticultural Crops
(274 ha)



Appendix 6. Report from the SMIS Stakeholder Workshop, June 26th 2017 10.00 – 14.30. Cranfield University.

1. Purpose of the workshop

- Introduce the SMIS project and current progress to key stakeholders;
- Receive feedback on what and how SMIS can deliver to the industry; and
- Demonstrate the potential benefits of SMIS to the industry.

2. Attendance

The original plan was to hold the SMIS Workshop in autumn 2017, targeting a large audience of key SMIS stakeholders. However, on advice from AHDB, the Workshop was brought forward to June 2017 to publicise the project as quickly as possible, now that the project's revised Milestones and Deliverables had been agreed between Cranfield and AHDB. It was also agreed that invitations would be focused mostly on SMIS participating industry partners (as listed in the original project proposal) and key growers' group representatives. Invitations (Appendix 1) were sent to 64 contacts on 19/05/17, explaining the purpose of the workshop, followed up (if needed) by 3 reminders (on 5th, 14th and 20th June).

The web based Event Brite was used to register workshop attendees. Many invitees told us they were unable to attend due to June being a very busy time of year. A total of 24 delegates attended.

Attendees

John Chinn	Cobrey Farms	john@cobrey.co.uk
Claire Donkin	West Growers	claire.donkin@westgrowers.co.uk
Mark White	PGRO	mark@pgro.org
Lizee Sagoo	ADAS	Lizzie.Sagoo@adas.co.uk
Guy Thallon	Farm Care Ltd	guy.thallon@farmcareltd.co.uk
Becky Howard	PGRO	becky@pgro.org
Coral Russell	Crop Associations Manager	Coral.Russell@britishgrowers.org
Cathryn Lambourne	AHDB	cathryn.lambourne@ahdb.org.uk
Steve Tones	AHDB	steve.tones@ahdb.org.uk
Bill Parker	AHDB	bill.parker@ahdb.org.uk
Gary Taylor	Hort Board	gary.taylor@ahdb.org.uk
Martin Evans	Field Veg Panel	mevans@freshgro.co.uk
Jim Dimmock	AHDB	jim.dimmock@ahdb.org.uk
Harley Stoddart	AHDB	harley.stoddart@ahdb.org.uk

Mike Storey	AHDB	mike.storey@ahdb.org.uk
Jerry Alford	Soil Association	jerry@littlebeacon.co.uk
Jane Rickson	Cranfield University	j.rickson@cranfield.ac.uk
Rob Simmons	Cranfield University	r.w.simmons@cranfield.ac.uk
Fady Mohareb	Cranfield University	f.mohareb@cranfield.ac.uk
Tom Kurowski	Cranfield University	t.j.kurowski@cranfield.ac.uk
Lynda Deeks	Cranfield University	l.k.deeks@cranfield.ac.uk
Jo Niziolomski	Cranfield University	j.c.niziolomski@cranfield.ac.uk
Caroline Keay	Cranfield University	c.keay@cranfield.ac.uk
Jason Carvalho	Cranfield University	j.carvalho@cranfield.ac.uk

3. Timetable

10.00	10.30	Arrival and refreshments	All
10.30	10.50	Welcome and Introduction to SMIS project	Jane
10.50	11.10	SMIS as a 'one stop shop' for diverse data / knowledge / information on soil management	Jo
11.10	11.40	Interactive session 1: Soil related data/ knowledge/ information Confidence in the sources? (score 1 [low] – 5 [high]; gives weighting in SMIS) How data/knowledge/information is used	All
11.40	11.50	Linking SMIS with LandIS® soil and environmental datasets	Caroline
11.50	12.10	Grower data, including use of proxy parameters, as direct metrics not currently measured (e.g. compaction)	Rob
12.10	12.40	SMIS system architecture	Fady / Tom
12.40	13.10	Interactive session 2: What's in it for farmers / growers / delegates…and how can they get involved?	All
13.10	14.00	Lunch, followed by tour of CHAP soil health assets (B54 and glasshouse)	All

4. Discussion of the presentations

Martin Evans, whose business focuses on carrot production, indicated that the project should have just identified 'horticultural soils' and then located key growers within these areas. Martin further indicated that SMIS needs to bring in data from field vegetable growers. He said the scope of the current datasets is too narrow and needs to be enlarged beyond the initial PGRO-based group to ensure as broad a representation of horticultural systems as possible. This can be achieved via communication with the various grower associations and contacts given to the project team by Martin, Cathryn Lambourne and others. (Subsequent to the meeting, RS contacted the following growers:

2. Worth Farms
3. Hay Farming
4. Jack Buck Farms
5. JSR
6. James Caley Farms
7. Albanwise Norfolk
8. Albanwise Yorkshire (x2)
9. Tompsett Growers
10. Sherwood Growers
11. Stawsons
12. Kettle Produce
13. Hardstaffs of Linby
14. Poskitts

+ grower groups via British Growers:

- Brassica Growers Assoc., [already presented SMIS at the spring 2017 meeting, but have built on this with another email].
- British Onion Producers
- Leek Growers Assoc.
- British Leafy Salads Outdoor Group
- Baby Leaf Growers Assoc.
- British Herb Growers Assoc.
- Outdoor Cucurbit Growers Association.
- Sweetcorn Growers

Martin also indicated that the Field Veg Panel (FVP) had not been fully consulted by AHDB during the commissioning of the project. This was frustrating as currently, 90% of the project funding comes from the Field Veg Panel with 10% from bulbs. The initial indication from AHDB was that the funding would be split across all the horticultural levy panels.

In addition, Martin indicated that there was also a general feeling of frustration amongst the FVP regarding the SMIS project as they had not received any project updates over the first 12-months.

John Chinn extoled the importance of the project and the huge potential of the project for new insights into soil management. He highlighted the need capture data/information relating to soil management, as sustainable soil management is a cross cutting theme across all AHDB sectors. There was general agreement amongst the participants to this view.

Over lunch-time discussions between Rob Simmons and Gary Taylor the importance of the SMIS project to the horticultural sector was discussed and confirmed.

There was also a discussion between Martin Evans and Fady Mohareb regarding how SMIS will be used as a tool to inform growers on best soil management practices; instead of the current approach of hiring a soil management expect with years of experience in this field. Fady explained that the purpose of SMIS is not to completely replace human experts, but rather to complement this, since 100% inspection of every single field is practically and economically impossible. SMIS will implement a model-based decision support system which incorporates data collected from as many farm fields and plots as possible. This decision support system will be made available to levy payers through AHDB Horticulture, and will simulate expert opinion on suggesting best practices by integrating data collected from the literature, experimental datasets, and growers' current practices. As the SMIS database grows in terms of size and diversity, the prediction accuracy of the implemented models will be enhanced. This allows advancement from macro-scale overviews towards more local, field-scale analyses for decision support. Furthermore, the model prediction will be accompanied by a confidence weighting measurement, depending on the number of evidences found in the database (i.e. fields, experiments and growers' data), supporting the "problem"- "cause(s)" and/or "problem"- "solution(s)" relationships.

Cathryn Lambourne during discussions on the way to B54 (tour of facilities) indicated that she would provide us a list grower associations/levy panels that we can target and make presentations to regarding SMIS as a basis for obtaining grower data (Appendix 2). A table of events where SMIS can be promoted is held on the shared drive at Cranfield and is regularly updated by the project team (Appendix 2).

Inclusion of LandIS data in SMIS hugely valuable to growers

The project team need to allow time to consider how/where SMIS should be funded/co-funded, hosted and developed beyond our initial contract – e.g. would one of AgriTech Centres (Agrimetrics/Agri-EPI/CHAP) provide the best fit? This will be scheduled as a separate meeting between AHDB and Cranfield project teams.

5. Interactive Sessions

The objective of these sessions was to

1. Identify all the sources of information used by stakeholders (e.g. growers / grower groups / AHDB staff / researchers /), and ensure these are captured within SMIS
2. Identify all soil management decisions considered by stakeholders and ensure they are captured by SMIS
3. Gauge the confidence in different sources of information available to stakeholders, using a scoring system from 5 (high confidence) to 1 (low confidence). This will inform the 'weighting' of the network visualisation [e.g. 'cause and effect' linkages] within SMIS.

Four breakout groups deliberately comprised of different stakeholders identified the following sources of information used by them, and the soil management decisions that were informed by these sources. In the second session, the level of confidence in each source was identified. The outputs from these sessions are summarised below.

Sources of information	Confidence score *	Soil management decision
Field visits / demonstrations		Destoning (limiting factor for carrots)
Crop walking		Soil nutrients / nutrition
Production planning		Soil pests and diseases (risks)
Individual experience / prior information of own farm		Timing of operations – start with harvest and work backwards e.g. tillage, planting, harvesting
Sampling and analysis e.g. soil moisture		Cover cropping
Rotational history		Type of tillage
Local knowledge		Soil erosion control
Farmer to farmer learning		Row length
Agronomist advice		Compaction prevention and alleviation
Technical notes / leaflets		Organic matter management
Press information		Pre planting cultivations
Case studies		Weed control
Peer reviewed papers		
EA		
Assurance schemes		
Experts (agronomists, researchers, ag. Chemical suppliers, academics,)		Not specified
Trusted (technical) advisors e.g. PGRO		
Weather forecast		
Anecdotal		
Farmers forum (internet)		
Technical meetings		
Commercial sources		

*Scoring categories:

5 (high confidence) 4 3 2 1 (low confidence)

Analysis of the outputs from the sessions concluded:

- Topography is important

- Cropping history especially on rented land (can be an issue to access)
- Soil type (i.e. texture)
- A range of sources of information were identified, but not necessary applied explicitly to soil management decisions.
    - Most are already covered in the categories of data / knowledge / information being captured in SMIS (i.e. research outcomes, literature (grey and peer reviewed), grower data, expert opinion)
- Anecdotal and individual experience very important in number of responses and confidence attributed to this – how can we capture this is SMIS?
- Role of the agronomist is important and inspires confidence
- Technical notes have a role to play, although confidence in them is variable
- Experts are respected
- Information is sought about specific activities: timing of operations (e.g. tillage), cover cropping, type of tillage, weed control.
- Other specific decisions (e.g. destoning, control of weeds, pests and diseases, were not linked to specific sources of information, although this does not necessarily mean it is not possible to do so. SMIS could play a role in making these connections more explicit.

6. Summary and concluding remarks

At first, there was some scepticism of the SMIS project from a few delegates, but the project team managed to reassure those present with our vision plus the project progress to date, including the grower data we have amassed and the tools being developed to interrogate the SMIS database. We also understood why Martin Evans and the Field Veg Panel in particular felt disengaged with the project. Rob Simmons met with Martin later in the week and Martin has put us in touch with a number of growers who we can approach for their farm data to feed into SMIS. Cathryn Lambourne has also offered to put us in touch with a number of grower associations and farmers groups to whom we can publicise the project and encourage them to be involved. (We already have a number of dates in the diary to speak to these groups).

The break out sessions were lively with valuable contributions from all delegates. There was very positive feedback from industry representatives that we were undertaking innovative approaches to use 'big data' from growers to inform better soil management.

Inevitably the future of SMIS after the end of the project (Nov 2018) was discussed, in terms of it being 'owned' by AHDB or by the growers / industry. This is an area for further discussions with the AHDB and Cranfield project teams.

Post script: Invitation sent to key industry partners and stakeholders.

Dear -----,

Soil management is at the heart of sustainable intensification. It has the potential to improve crop yield (both quantity and quality), whilst protecting the soil and water resources that underpin horticultural agribusinesses today. However, sources of information and guidance on soil management in horticulture are currently unstructured, uncentralised and difficult to find or access.

AHDB Horticulture have commissioned a research project (CP107d) to develop a Horticultural Soil Management Information System (SMIS), that will explore how better use of data, information and knowledge on soil management problems and practices can help growers, agronomists and land managers make decisions that maximise crop production without damaging their key farm asset – the soil.

We are holding a key stakeholder workshop on Monday, 26th June at Cranfield (10 am to 2 pm) to introduce the project, to receive your feedback on what you want SMIS to do for you, and to demonstrate the potential benefits of SMIS to the industry. There will also be a tour of Cranfield soil research facilities after lunch for those that can stay.

We appreciate this is a very busy time of year, but do hope you can make it. Please register for the event by Friday, June 16th at the following link:

https://www.eventbrite.com/e/horticultural-soil-management-information-system-tickets-34655208662. Please click on Register and then enter details as instructed.

Morning refreshments and lunch are included and we are happy to meet reasonable travel costs for attendance.

If you have any questions about the workshop or the SMIS project, please do not hesitate to contact me.

With best wishes

Jane Rickson
Professor of Soil Erosion and Conservation
School of Water, Energy and Environment
Building 52A, Cranfield University, Cranfield, Bedfordshire MK43 0AL
E: j.rickson@cranfield.ac.uk
T: +44 (0) 1234 750111 x2705
W: www.cranfield.ac.uk

Appendix 7. Events at which the SMIS project has been / will be promoted (as at 31/10/17)

Event	Date	Attended by….	Number of delegates	SMIS project mentioned	Invited to offer farm or field data
British Herb Trade Association field event (Millets Farm, Oxfordshire) 'Soil in Horticulture'	25th Aug 2016	Rob and Lynda	40	✓	
British Carrot Growers Association - variety day and trade exhibition (West Knapton, Yorkshire) 'Soil in Horticulture'	6th Oct 2016	Lynda	100	✓	✓
20 Club Farmers (Stamford, Lincs.) 'The importance of soils for the sustainable intensification of agriculture'	10th Nov 2016	Jane	30	✓	✓
GREATsoils workshop: Soil Health and Farm Viability (Lichfield) 'Keeping soil in the field - managing runoff and erosion in row crops'	22nd Nov 2016	Lynda	10	✓	✓
Carrot Expert Group Meeting, (Croxton Park, Thetford, Norfolk) 'Understanding the biotics and opportunities to improve soil health'	24th Nov 2016	Lynda and Jane	12	✓	✓
Vegetable Consultants Association Annual Conference (Stilton, Peterborough) 'Challenges of Sustainable Soil Management': Metrics of soil health'	29th Nov 2016	Rob and Lynda	25	✓	✓
Hutchinsons Annual Conference Whittlebury Hall, Northants, 'Managing soil health'	2nd Feb 2017	Rob	240	✓	✓
AHDB Legume R&D Panel	10th Feb 2017	Rob	15	✓	✓
HLH Vegetable Conference (Lincolnshire) 'Soil management in vegetable production'	21st Feb 2017	Lynda	60	✓	✓
Nottinghamshire Farmers Business Association (Nottingham) 'What have we done to our soils over the last 40 years…. And looking to the future, what do we need to do?'	27th Feb 2017	Jane	70	✓	✓
International Brassica Conference (Lund, Sweden)	8th Mar 2017	Jane	70	✓	✓

Event	Date	Attended by….	Number of delegates	SMIS project mentioned	Invited to offer farm or field data
'The importance of soil health in brassica production'					
Brassica Growers Association R&D Panel	16th March 2017	Rob	8 attendees including Cathryn Lambourne of AHDB.	✓	✓
Meeting with HMC Peas Growers	20th March 2017	Rob	5	✓	Additional grower group manager contacts provided
Prot. Ornamentals, Bulbs & Outdoor Flowers contact Georgina Key georgina.key@ahdb.org.uk	20th April 2017 AHDB Stoneleigh				
SMIS Stakeholders meeting Growers, AHDB staff, researchers, industry etc. etc.	June 26th 2017 Cranfield	All project team	30	✓	✓
British Carrot Growers Association (BCGA) (Jane accepted invite from Coral Russell)	29th June 2017 Hexgreave Hall, Newark	Rob	18 attendees including Cathryn Lambourne of AHDB	✓	Supported by Martin Evans, Growers agreed to provide data
Hardy Nursery Stock contact Georgina Key georgina.key@ahdb.org.uk ✓ contacted 13/04/17	29th June 2017 (Please note change of date – was 9th June) AHDB Stoneleigh (tbc)				
Tesco Soil Management Meeting (Taylorgrown, Houghton Hall) 'Achieving soil health through collaboration	25th July 2017	Lynda	30	✓	✓
Frontier Technical Team	7th September 2017	Jane / Rob	12	✓	

Event	Date	Attended by….	Number of delegates	SMIS project mentioned	Invited to offer farm or field data
British Herbs Summer event (at Tozer Seeds, Stoke D'Abernon, Surrey) (Rob has agreed to do a field demonstration	14th September 2017	Rob	35	✓	✓
AgriTech East SMART farming event, Cranfield University	14th September 2017	Jane	100	✓	✓
Carrot Grower Data Collection [Strawsons Ian Holmes] Discussions relating to soil/crop management issues in Carrots	26th September 2017	Rob	2	✓	Data Collection
British Carrot Growers Association (BCGA) 2017 Variety Demonstration (Rob and Lynda Invited to attend and give field demonstration on 'Soil Management'	5th October 2017	Rob and Lynda	173	✓	Data offered
Visit to Watts Farms (Kent) Tomasz Zdunek (Farm Operations Manager)	10/10/17	Rob			
AHDB Legumes R&D Panel Update on SMIS provided	19/10/17	Rob	12	✓	✓
Field Vegetable Panel contact Cathryn Lambourne cathryn.lambourne@ahdb.org.uk	9/11/17 Woodside Hotel, Kenilworth	Rob and Jane			
UK Vining Pea and Bean conference (invite from Coral Russell) Theme this year is "Health and Nutrition" from Crop to Consumer. Within the Crop Nutrition and Health I would like a focus on Soils; I was wondering if I could ask you about your availability to firstly attend the event but also to present AHDB Hort SMIS at the event. I also wanted to make sure that the SMIS research project is topical to Vining Pea and bean growers? The conference is on the 21st November at the Kingsgate centre, Peterborough. We are looking to attract growers, managers, field staff, agricultural managers and senior managers	21st November 2017	Rob			

Event	Date	Attended by….	Number of delegates	SMIS project mentioned	Invited to offer farm or field data
from processors and freezing companies to the industry event.					
Tree fruit contact Rachel Lockley rachel.lockley@ahdb.org.uk ✓ emailed 13/04/17	28/11/17 NIAB-EMR				
British Herb Trade Association Jane accepted invite from Coral Russell, Crop Associations Manager 13/04/17	30/11/17 (Cranfield)	Rob and Jane			
Soft Fruit contact Rachel Lockley rachel.lockley@ahdb.org.uk ✓ emailed 13/04/17	6 Dec 2017 Venue to be arranged				

Appendix 8. The Grower article (submitted to AHDB, 15/09/2017)

AHDB Grower 2017 – Oct/Nov

GREATsoils – CP107d SMIS

<Headline>

Big Data, Big Knowledge

<sub header>

Jane Rickson, Rob Simmons and Lynda Deeks from Cranfield University take a holistic cross-disciplinary approach to mining and using soil management data.

<body text>

Many growers already collect data on aspects of crop agronomy, field operations and aspects of soil health as part of their routine farm management. While some of it is used for business planning or to support assurance and certification schemes, there is underutilised potential that could be used to optimise benefits on farm.

Some of this data has the potential to enhance the productivity and competitiveness of growers' businesses, including data that could support sustainable soil management or drive innovation in cropping systems. However, these potential benefits can't be realised from data from one business on its own or even a few businesses working together.

To get the best out of this data means combining it with as many other coherent data sets as possible, generating what's known in information technology jargon as 'big data'. Storing, cleaning, transforming and analysing such large and complex collections of different kinds of data is beyond the normal computing capacity of most individual businesses. Meaningful interpretation of such datasets requires new and emerging methods of data management and processing known as 'agri-informatics' to unearth valuable insights that would otherwise have remained hidden. These insights can then be presented to growers, agronomists, advisers and researchers.

One area where big data has real potential is in helping to improve soil management in horticulture, where intensive production practices can lead to soil health issues.

Many of the problems caused by these issues were identified by growers in a survey carried out by AHDB in 2013 as part of a project analysing soil management research and knowledge transfer in horticulture (CP 107). The project confirmed how important soil management is to crop productivity, not least because of soil borne pests and diseases, and degradation resulting from compaction, erosion, loss of organic matter and loss of biodiversity.

The review revealed that more work was needed on areas such as:

- Understanding what soil parameters need to be measured by growers, and developing 'smart technology' to measure these properties
- Making the best use of soils data
- Fine-tuning nutrient management
- Better systems to manage intractable soil borne diseases
- Using 'precision agriculture' to improve soil management
- Using and getting the best from soil amendments and other inputs
- Managing soils for cropping consistency

<heading> The emergence of an information system

The need for a system to enable the industry to make better use of existing whole farm, cross-rotational data is being addressed in the project CP107d (which makes up a third of the GREATsoils programme of soil-related work), the development of a horticultural soil management information system or 'SMIS'. It's exploring how better use of data and information on, and knowledge of, soil management practices can help growers make informed decisions that achieve the best possible yield and quality from the crop without compromising what is their key farm asset – the soil.

SMIS has already begun to collate relevant data, information, knowledge and expertise on sustainable soil management from many diverse sources: scientific and technical literature, expert knowledge, anonymised data collected by growers, outputs from AHDB research projects (including those in the CP107 programme) and anecdotal information from growers and crop sector groups.

This integration of wide-ranging data forms and sources makes SMIS unique – such a holistic approach has never been attempted before. The pool of knowledge and information that's being gathered will form a 'rule-base' on the causes of soil management issues and how soil management practices can solve them. This rule base will encompass the management of compaction, erosion, organic matter content and key soil borne diseases. The database will identify particular soil and crop management factors affecting yield and yield quality in a rotational context. Obtaining coherent datasets that encompass full rotations is critical to identify positive or negative effects of particular operations/practices on yield, yield quality and to promote sustainable soil management.

Growers' data is an essential element of SMIS because it brings an understanding of the timing, type and frequency of farm operations in a rotational context, including those that can lead to soil degradation, and provide an evidence base for those practices that promote sustainable soil management. Most importantly, anyone who agrees to share their data through SMIS can be confident that it will be fully anonymised and used solely to generate the SMIS 'rule-base'.

The value of this whole farm data is enhanced further by linking it to soil data already held in Cranfield University's LandIS database (landis.org.uk) and to meteorological data sets. This will make it possible, for example, to define the environmental conditions when land can be worked without causing soil compaction – known as 'workability' days – or how quickly soil that has suffered compaction can recover naturally.

Once the database has been compiled, computer-based 'agri-informatics' analytical tools known as 'data-mining' techniques will be used to interpret the data to tease out any common patterns. These may relate to a particular soil type, crop, rotation, geographical location or a combination of these factors. Such knowledge will help growers with similar conditions anywhere else in the country.

<heading> Benefits for levy payers

SMIS will collate, harmonise and integrate information and data on soil management that is currently spread across tens or even hundreds of individual organisations. SMIS will provide a wealth of relevant information on sustainable soil management, specifically related to horticulture, from one central source.

For individual growers, patterns in their own farm data are often obscured by the variations in soil management practices and their effects from season to season, year to year and from field to field. This 'noise' starts to fade as the pooled dataset gets bigger. Patterns that aren't visible in an individual data set are more likely to be revealed and can be used, for example, as the basis for best practice guidance on soil management.

Additional benefit can also be drawn from the grower data by looking at the full rotational context and the operations associated with it. For example, we may be able to identify the optimal rotations and farming practices to mitigate soil borne pests and diseases and so help the industry to reduce reliance on chemical crop protection products.

As more and more farmers and growers contribute their farm data, the rule-base will become more robust and the resulting guidance on soil management more reliable. As a result, growers will be able to better maintain soil health holistically by improving physical, biological and chemical soil properties to ensure their soils are resilient. A healthy soil is better able to receive, retain and release water, provide good support for farm traffic, resist erosion by wind and water, and support a diverse microbiology that promotes nutrient cycling, structural re-generation and resilience to soil borne disease. Healthy soils, in association with appropriate tillage and agronomic practices, promote good seed germination, crop establishment and optimise yield and yield quality.

SMIS is very much a live and dynamic project: new data, knowledge and information is continually being added. As the database expands, SMIS's explanatory and predictive capability increases. Our ability to interpret the database will be strengthened over time, as new and more sophisticated statistical and data-mining techniques emerge.

For further information on SMIS and details of how to contribute your own farm data, please contact Lynda Deeks at l.k.deeks@cranfield.

<Box out>

CP 107d Development of a horticultural soil management information system (SMIS)

Term: November 2015 to October 2018

Project leader: Jane Rickson

Research consortium: Cranfield University, James Hutton Institute and PGRO

<Box-out> Get involved!

Why not join us at one of the following upcoming workshops to find out more?

Events to go in here…

Discover more online **www.ahdb.org.uk/greatsoils**

<Pull quotes>

"For individual growers, patterns in their own farm data can be obscured by variety of soil management practices"

--END--

Word count: 1190 (inc. box outs).

Appendix 9. SMIS DATABASE ARCHITECTURE TECHNICAL DOCUMENTATION V1.0

Revision History

Version	Description	Date	Author
1.0	Original database architecture description. Documentation is largely exhaustive for grower data and literature data and reflects work done. For experimental data, the documentation remains a general development specification in this version.	22/08/2017	Tomasz Kurowski

Abbreviations

API Application Programming Interface

DBMS Database Management System

HTTP(S) Hypertext Transfer Protocol (Secure)

JSON JavaScript Object Notation

MEAN MongoDB, Express.js, AngularJS, and Node.js

NG National Grid

OS Ordnance Survey

REST Representational State Transfer

SMIS Soil Management Information System

SubVESS Subsoil Visual Evaluation of Soil Structure

VESS Visual Evaluation of Soil Structure

VSA Visual Soil Assessment

XML Extensible Markup Language

1. Introduction

1.1. Purpose

The following document describes the design of the database back-end, which forms a vital element of the Analytics Toolkit developed as part of the Soil Management Information System (SMIS) project. The purpose of the document is to provide an overview of the design, intended to serve as an implementation guide for the developer and as an accurate description of the technical details of the system accessible to the end user.

It should be noted that as SMIS remains in active development, this document will be updated as the implementation progresses until the hand-over of the system (November 2018) in order to ensure its thoroughness and accuracy.

1.2. Scope

The collection, storage, manipulation and interrogation of information obtained from diverse data sources related to the effects of soil management practices on horticultural crop productivity and environmental protection are central objectives of the SMIS project. Therefore, the specific database management solutions to be used are of primary importance to the project's success.

The document presents the specific database design decisions taken, alongside the technical considerations which guided them. In particular, these include how and by whom the system is to be managed and maintained, how the database will be populated throughout the course of this project, and how the data will be accessed.

Thus, alongside a technical description of the database implementation in the narrow sense, the document also discusses the system context. This includes both the immediate “upstream” interface of the database, that is the types of data used and the ways in which they are parsed (i.e. processed and inserted into the database), as well as the immediate “downstream” interface of the database, that is the Application Programming Interface (API) which will provide the means of interrogating and manipulating the database by the SMIS Analytics Toolkit.

Details of the data collection and SMIS Analytics Toolkit functionality are not discussed here, except in how they directly impact the designs of the parsing tools and API respectively.

2. System overview

The SMIS database back-end uses the MongoDB database management system (DBMS) and forms a part of a larger structure. It is responsible for storing and allowing the use of three primary types of data:

Grower data

Experimental data

Literature data.

The end users interact with the data exclusively through a web application (the SMIS Analytics Toolkit) which in turn accesses the database through a representational state transfer (REST) API implemented using Mongoose, Express.js and Node.js.

The database is populated with curated data, processed by parsing scripts implemented in Python. For performance reasons, these scripts access the database directly (using the *pymongo* library) rather than through the REST API. The scripts are considered part of the SMIS system although they are not accessible through the web application. This is discussed more in-depth in section 7.1.

As adding new datasets is intended to be the task of a curator / administrator rather than a standard web application functionality, the data volatility is assumed to be low for most use cases.

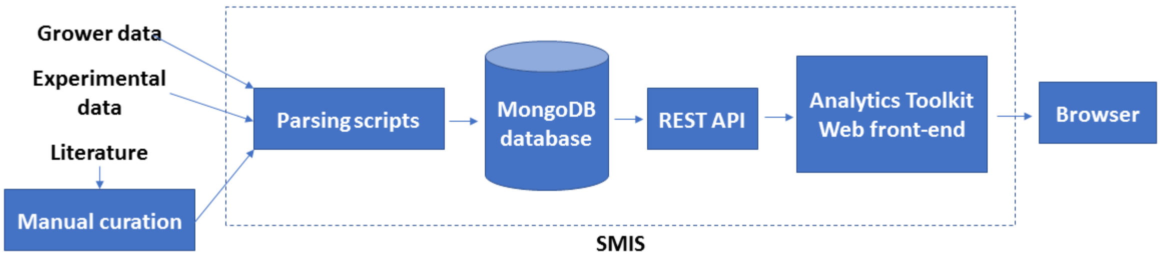


Figure 19: Information flow in SMIS

SMIS is intended for internal, local use by an organisation (i.e AHDB) rather than Internet-wide public access. This means the traffic and demand on concurrent access are assumed to be relatively low.

2.1. Choice of DBMS

MongoDB 3.4 was chosen as the database management system (DBMS) used in the SMIS back-end. The primary reason for this choice over a relational database was the schema-less nature of MongoDB. This is a crucial advantage as SMIS integrates data from disparate sources and the number of supported data formats is intended to be expandable beyond those represented among the datasets available during development. Any added formats may include new data fields and types of data, which are trivial to add to a document-oriented database like MongoDB. In a relational database expanding the capabilities in an equivalent manner could require significant changes to the database schema and possibly the rest of the application.

Relational databases hold a significant advantage in how they strictly enforce data integrity and validity. However, in SMIS the main threat to data integrity comes from inconsistencies caused by human error (e.g. typos) or differences in data entry practices between data providers (grower groups). This shifts the onus of ensuring data integrity and validity to the parsing stage, outside the control of the database management system, and largely negates the advantage of relational databases.

An additional advantage of MongoDB is its ease of integration with web applications through its inclusion as part of the MEAN stack, a popular bundle of technologies combining the database management system with the Express.js and AngularJS frameworks running on a Node.js server.

2.2. System requirements

Dedicated[1] PC running a modern Linux, macOS, or Windows Operating System

At least 10 GiB of storage space

Network access (HTTP/HTTPS port)[2]

While the SMIS development environment is Unix-based, all the technologies used are also available on Windows. The application will be tested in that environment to ensure it remains fully functional. System-specific installation scripts will be prepared. Alternatively, a containerised/virtualised solution, such as a Virtual Machine including a working system, could also be provided.

The database back-end in itself does not require a network connection, as it is only intended to be accessed by the SMIS Analysis Toolkit through a local port (on some systems this may require modifying firewall settings). However, the SMIS web application which depends on the database does require network access so that it can be accessed from other machines.

The currently gathered data (as stored in the largely unnormalised and highly indexed database) take up less than 5 GiB of storage space and are not expected to grow much larger than that threshold before the end of the project (November 2018). Storage is therefore unlikely to be a problem, although further data gathering beyond the end of the project could raise the data volume to arbitrarily large sizes.

[1] As the SMIS Analytics Toolkit is meant to be accessed over a local network, the server machine must run continuously.

[2] This requirement concerns the SMIS Analytics Toolkit as a whole rather than the database back-end itself.

While traffic requirements are assumed to be relatively low, the technologies used are known to scale well. A series of full-scale performance tests will be carried out before the end of the project. These will allow for an accurate assessment of the scalability and requirements of the SMIS system if the database was to grow much larger or if the traffic was to increase significantly.

2.3. Grower data

Gatekeeper format

Gatekeeper is an agronomic record-keeping solution developed by Farmplan and used by some growers in the United Kingdom. All grower datasets used in the development of SMIS are Gatekeeper datasets, manually exported from the application according to export instructions provided to growers (see accompanying document: Protocol for extracting data from Gatekeeper).

The default data export format used by Gatekeeper is a “Gatekeeper XML” file, though the same data can also be saved as an Excel spreadsheet. The two formats are in fact largely equivalent, as the XML output is a “flat” XML file, with unnormalised data stored between non-nesting tags representing individual rows of a spreadsheet.

Gatekeeper synchronisation

Besides its use as a stand-alone desktop application, Gatekeeper also supports a client-server paradigm in which a user can synchronise the contents of their local copy of the application with a centralised server, or even share it with others (e.g. an agronomist) by providing a special access key.

While this suggests the possibility of directly integrating SMIS with Gatekeeper without the need for intermediary XML / spreadsheet files, this was not attempted, as no public API for emulating this client-server integration is available. Additionally, such close integration with a proprietary software product would be subject to obsolescence as the product changes, presenting a difficult maintenance challenge.

Other formats

Agronomic record-keeping solutions other than Gatekeeper (such as Muddy Boots) also exist and maintain a significant share of users among growers, but they are not represented among the datasets used during SMIS development and initial testing as all growers’ data collected to date are in Gatekeeper format. As SMIS is intended to be extendible and capable of being adapted to handle disparate sources of grower data, avoiding overly tight integration with the Gatekeeper format and its conventions is among the objectives of the back-end design.

Schema

Information derived from grower datasets is stored across three collections as shown in Figure 20.

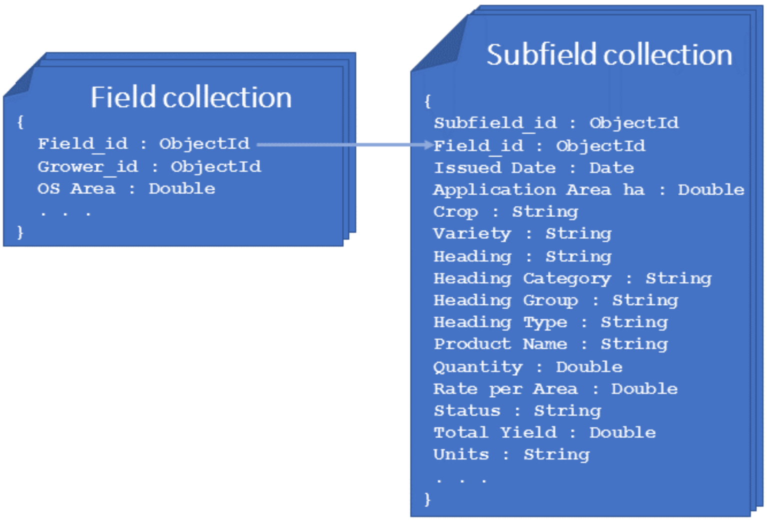


Figure 20: Grower data collections diagram

The **Field** collection stores farm field–level information such as a farm field's unique identifier; the unique identifier associated with the grower who supplied the dataset; the Ordnance Survey (OS) area of the farm field; and potentially field data originating from external databases. It should be noted that grower data in SMIS is subject to anonymisation as described in section 3.1.1, so both the farm field and grower identifiers are generated at the time of parsing by (respectively) hashing and tokenisation.

The **Subfield** collection stores information about any farm field operations, cropping, yields associated with a specific area within a farm field identified by its unique identifier (corresponding to that stored in the Field collection). This is a necessary construct as sometimes farm field operations and crop rotations are not applied to entire farm fields (as defined in grower datasets), but to smaller subdivisions. An explanation of how this information is captured is provided in section 3.1.1. The Subfield collection is indexed on every single data field to enable for efficient execution of complex query and server-side pagination of the results for the intended use downstream (i.e. in the SMIS Analytics Toolkit).

The third collection derived from grower datasets is the **Field_Vocabulary** collection, which serves an auxiliary function in parsing grower datasets as described in section 3.1.1. It contains updateable arrays of "known" values for each data field (column) in the Subfield collection, which make it possible to detect unexpected (erroneous or novel) values in datasets during supervised parsing. Those known values are paired with "Canon" values, allowing for normalisation (or canonicalization) of data field values.

2.4. Experimental data

The primary experimental dataset used during SMIS development came from the "CP107c – Soils Programme: The application of precision farming technologies to drive sustainable intensification in horticulture cropping systems" project. The data are collected in an Excel spreadsheet, with the results of different soil structural analyses spread over multiple sheets (one per analysis type) and identifiable by shared farm field identifiers and data collection dates, in effect partially normalised.

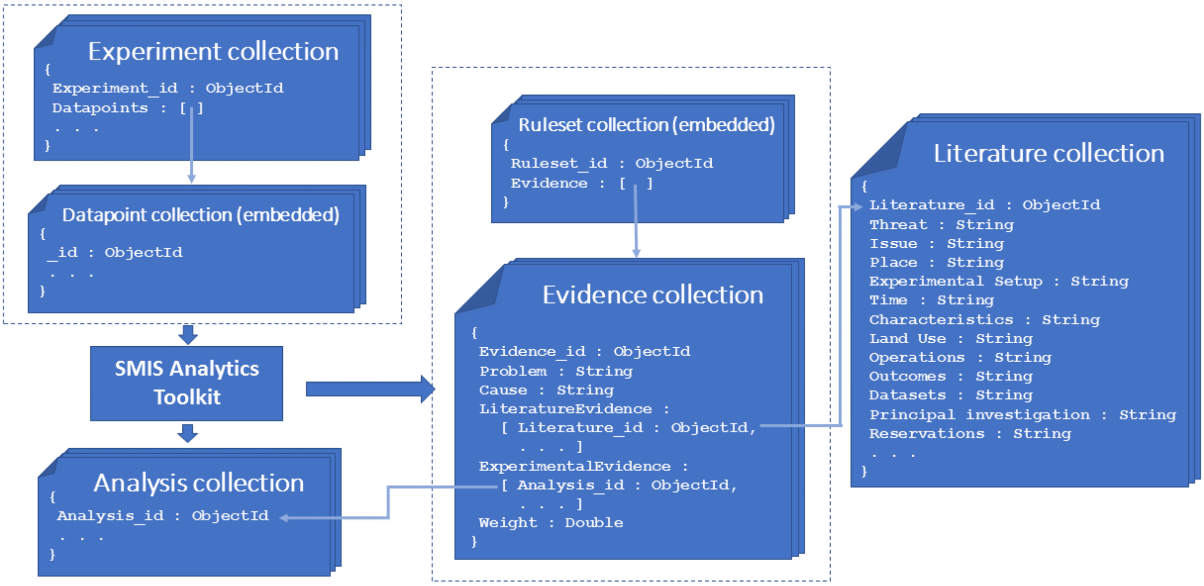


Figure 21: Diagram of the database schema and relationships for Experimental, Literature, Analysis and Evidence data

A general overview of the Experiment collection schema is shown in Figure 21. Experimental datasets are uniquely identified by a generated token and contain an embedded collection of schema-less data points which store measurements and other metadata. Each experiment entry may also be annotated with metadata. This very generic structure contains few constraints and thus allows for future inclusion of experimental datasets other than the ones used in development.

Within SMIS, all the CP107c data are grouped in a single experimental dataset. Results from different types of analyses are stored together in the embedded Datapoint collection, grouped (denormalised) by the farm field and date of analysis. The types of results stored include:

Soil bulk density measurements

Soil penetrometer scores

Subsoil Visual Evaluation of Soil Structure (SubVESS) scores

Topsoil analysis

Visual Evaluation of Soil Structure (VESS) scores

Visual Soil Assessment (VSA) scores

The interpretation of experimental datasets is undertaken in the SMIS Analytics Toolkit downstream.

2.5. Literature data

The literature data used in the SMIS project was manually curated based on available sources to establish connections between soil degradation threats, soil management practices and other factors, which together would comprise a unique and useful overview of the current state of the art on those relationships. Summaries of the literature sources were manually generated and are stored in spreadsheets, which are then imported into the SMIS database and viewable in the Analytics Toolkit. The specific data fields for each item of data are listed in the "Literature collection" part of the schema diagram in Figure 21.

2.6. Analysis data

The Analysis collection is a simple schema-less document collection used to store the results of analyses and visualisations generated through the SMIS Analysis Toolkit web application. As the results of these analyses and visualisations can be very disparate and the specific analysis use cases and tools are still under development, the collection remains schema-less other than associating a title and unique identifier with the arbitrary analysis output and possible metadata.

2.7. Evidence data

The Evidence dataset is a derivative collection based on Literature and Experiment data (via Analysis results). Rows of the collection represent "rule bases" linking causes and effects identified either by manual literature curation (see section 2.5) or by the results of analyses based on experimental or grower data (see section 2.6).

3. System context

3.1. Data parsing

This section is intended to provide a general overview of the SMIS functionalities immediately upstream of the database, i.e. the parsing pipelines which populate the MongoDB collections based on the datasets gathered during the project. The pipelines are all implemented in Python and rely on the Mongoose API to connect to the MongoDB database, as this allows for more efficient bulk insertion of data than what could be achieved using the REST API used downstream by the SMIS Analysis Toolkit to access the database contents.

Vocabulary normalisation

As noted in section 2.1, inconsistencies in the input datasets used by SMIS are a major issue which needs to be addressed at the parsing stage. This is particularly true for the Grower data, which is generally entered manually and managed by individual growers. Inconsistencies can stem from human error (e.g. typos) or from data entry conventions (e.g. particular nomenclature or abbreviations) which may be fully consistent within a single organisation, but often vary between them. For example, one grower dataset might use the term "Vining Peas" while another uses "V.

Peas" to refer to the same crop. For SMIS to interpret the data correctly, these conventions need to be reconciled. The same problems can also be encountered in the experimental datasets.

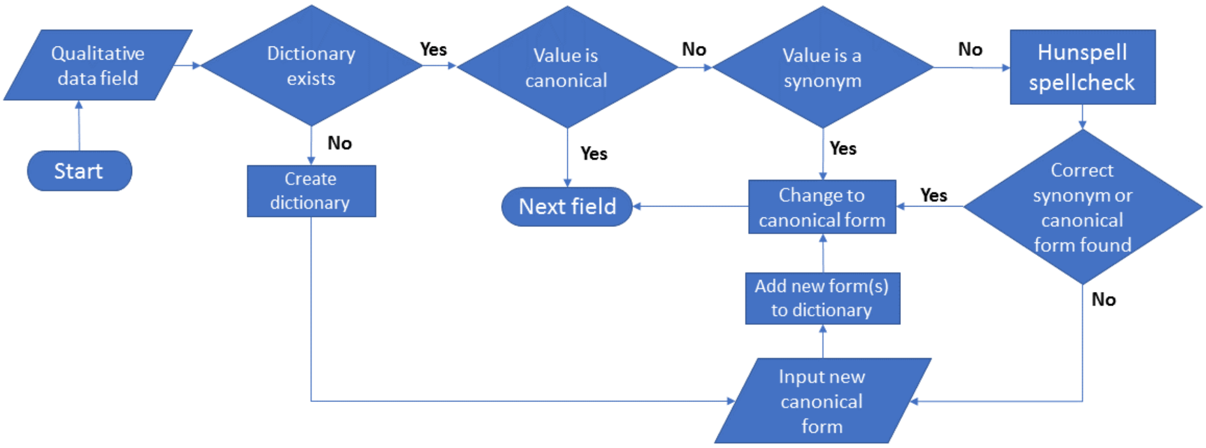


Figure 22: Vocabulary normalisation process for qualitative data fields

The Vocabulary Normalisation parser module is designed to resolve both issues by storing "canonical" values which represent the "correct" version of each term used by SMIS. Whenever a novel value is encountered, the parsing tool attempts to either: a) if it represents a known entity, correct it to a pre-existing (or new) canon value or b) if it represents a genuinely new entity, add it to the known data field vocabulary as a canon value or a synonym to a canon value. An overview of this process is shown in Figure 22. It should be noted that vocabulary normalisation is only applied to data fields containing qualitative (textual) information. Data fields containing numerical values do not undergo normalisation.

Using the stored vocabulary as a dictionary, the widely used, open-source Hunspell spell-checking library is used to propose alternative (canonical) spellings or forms to the user who can then either select any of the proposed versions, view the entire vocabulary list for a data field to choose a different one, or enter a new canonical form into the database.

It is also possible to export and import vocabularies in tabular form. This functionality was developed primarily to speed up grower and experimental data parsing during development and testing, but the import option can also be used to effectively "front-load" the SMIS database, populating it with known values (e.g. herbicide names) which would ensure the validity of the terms used, while simplifying the work of a human curator during parsing.

3.1.1. Grower data

The parsing pipeline for grower data can be seen in Figure 23. The individual steps are discussed in sections below.

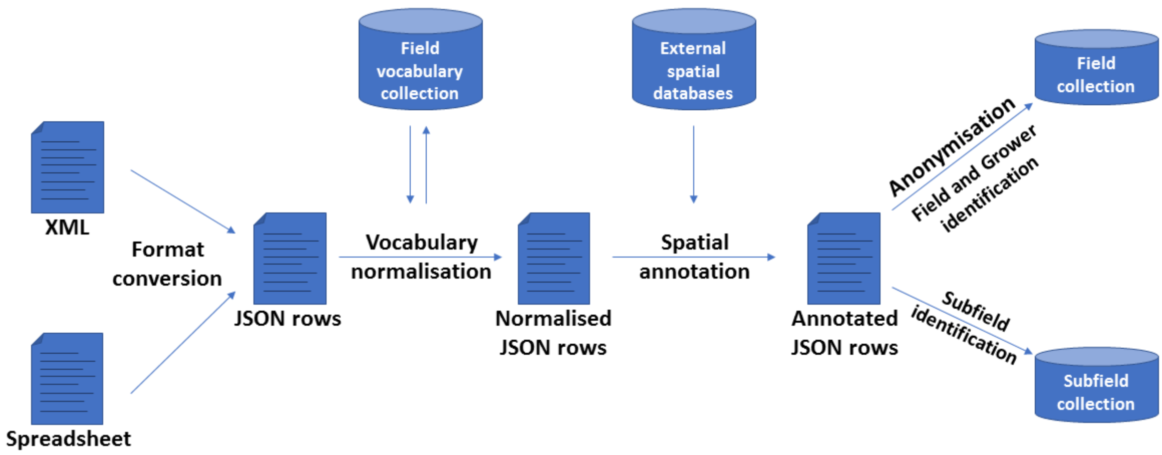


Figure 23: Grower data parsing pipeline

Format conversion

As mentioned in section 2.3, the currently handled formats are the Gatekeeper XML format and Excel/CSV spreadsheets. Both formats are unnormalised and are thus straightforwardly convertible into sets of equally unnormalised JSON objects, each representing a single row of data. These JSON objects form a single canonical entry point for pipeline input, and therefore expanding SMIS to handle a novel grower data format requires implementing a script to convert that format into the JSON representation. For normalised datasets, this would require explicit denormalisation.

Vocabulary normalisation

JSON-formatted Grower data undergoes vocabulary normalisation as described in section 0.

Spatial annotation

As data stored in the database is intended to be anonymised, identifiable spatial information such as Ordnance Survey Map Sheet identifiers and National Grid (NG) codes is not stored. Any data from external databases which depends on knowing this spatial information, such as LandIS-derived soil information, need to be fetched and stored at this stage of the pipeline, before anonymisation. Database-specific import scripts can be included to pull such data on a farm field by farm field basis.

Anonymisation

Before being discarded to anonymise grower data, the Map Sheet identifiers, NG codes, and farm field names are concatenated and processed by a SHA-1 hash function to generate unique identifiers for each farm field.

An alternative to this approach would have been tokenisation, i.e. the generation of novel, random identifiers. While this could have been more secure, it would also prevent the possibility of importing updates to existing datasets (newer or previously missing data field information), as a relationship between the old and new data could not be established. It should be noted that tokenisation is used for identifying sub-fields and growers as described in the sections that follow.

Grower and FARM Field identification

Each imported dataset is assumed to come from a single grower. However, successive (or otherwise separate) datasets can originate from a grower whose data are already in the database. While any duplicate entries can be discarded, novel data (e.g. covering later dates, or including data fields which were not previously available) has to be connected with existing data in order to give a full picture of factors such as the rotational context, effectively updating the database.

After the farm field anonymisation step described in the previous section, the identifier of each farm field in the dataset being imported is compared with those already present in the Field collection. One of the following three options is taken depending on that comparison:

If any of the farm fields being imported are already present in the database, the whole dataset is interpreted as coming from the same grower and any new farm fields are inserted into the Field collection alongside farm field–level information (e.g. “OS Area” and the unique grower identifier).

If none of the farm fields in the dataset being imported are present in the database, all of them are inserted into the Field collection alongside farm field–level information including a new, randomly generated grower identifier.

If the dataset being imported includes farm fields marked as originating from different growers (which is not the case in any of the data used in development, but is theoretically possible), the parser requires the user to choose which novel farm fields should be grouped with which group of existing farm fields. The new farm fields are then inserted into the database alongside farm field–level information.

Note that individual growers themselves are not represented by a separate collection in the database and are only represented by the randomly generated identifiers used in the Field collection.

Subfield identification/inference

In practice and in most grower datasets, farm fields are not indivisible entities, and multiple crops (accompanied by multiple corresponding sets of operations) may be grown concurrently on a single farm field, creating separate rotational contexts, separate histories of field operations, etc. For the purposes of SMIS, this separation of farm fields into virtual “sub-fields” needs to be captured.

This information may be contained in Gatekeeper data (“Part Field Reference” data field), but it is missing from many real-life datasets, including some of those used in SMIS development. It is therefore necessary for the parsing pipeline to attempt to *infer* the subfields by their shared area.

A customisable hectarage margin of error (10% by default) is used in subfield inference, as the recorded areas of operations are rarely exactly the same.

Data rows for each subfield (identified either by the Gatekeeper “Part Field Reference” or by the subfield inference functionality) are inserted into the Subfields collection together with a farm field identifier (allowing for aggregation with the Fields collection) and a new, randomly-generated unique subfield identifier. Individual operations covering the entire farm field rather than the area of a subfield are duplicated for each of the subfields.

3.1.2. Experimental data

The parsing pipeline for experimental data can be seen in Figure 24. The individual steps are discussed in sections below.

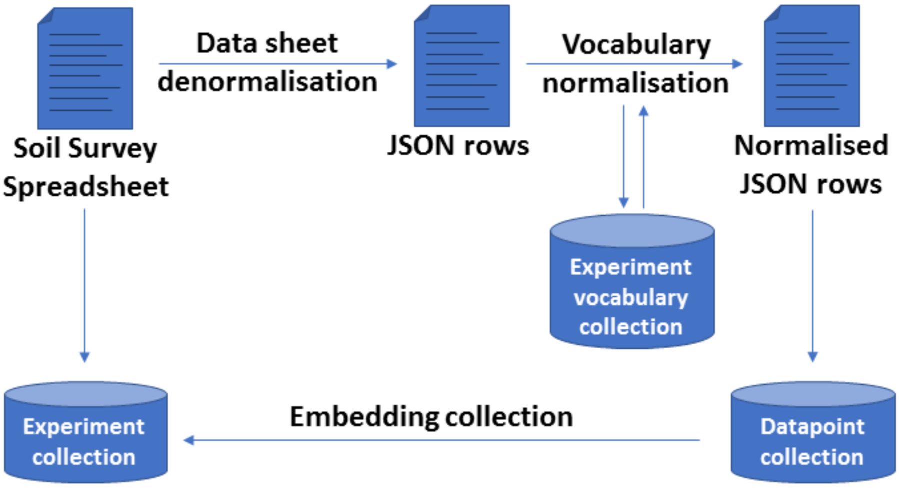


Figure 24: Experimental data parsing pipeline

Data sheet denormalisation

The “CP107c – Soils Programme: The application of precision farming technologies to drive sustainable intensification in horticulture cropping systems” Excel spreadsheet data points collected for the same farm fields on the same dates are split over multiple sheets, with one sheet per method. A Python script is used to bring the data from those sheets together, based on the collection dates and farm field identifiers. The collected data rows are converted into a JSON format, ready for insertion into a MongoDB collection.

When other experimental datasets (such as data from the “FV 447 - Carrots & Parsnips - Developing a strategy to control Free Living Nematodes” and “FV 380 - Identifying critical soil P in vining pea crops” projects) will be added to the SMIS database, novel scripts tailored to those datasets will have to be implemented for this stage of the data parsing pipeline.

Vocabulary normalisation

JSON-formatted experimental data undergoes vocabulary normalisation as described in section 0. In contrast with grower data, inconsistencies resulting in differences in terminology are not a major issue, but other sources of inconsistency such as typos do require correction.

Data storage

Finally, the collection of denormalised datapoint rows, with all of their quantitative data fields converted to canonical forms, are embedded in an element stored in the Experiment collection, identified by a randomly generated identifier.

3.1.3. Literature data

Literature data is subject to manual curation as part of the SMIS project and summaries for all sources are prepared in the form of spreadsheets in a format directly mirroring the one shown in section 2.5. As such, parsing the data consists only of a single step which converts the spreadsheet into JSON format and persists the rows into the Literature collection. As with the experimental data, this conversion process is conducted using a specialised Python script specific to the formatting used in data preparation rather than through a more generic tool.

Appendix 10. SMIS USE CASE DOCUMENTATION V1.0

Revision History

Version	Description	Date	Author
1.0	Original SMIS use case report. “Established query” (see section 0) use cases will be updated as their viability is tested and their interfaces are implemented.	9/10/2017	Tomasz Kurowski

Abbreviations

API Application Programming Interface

NG National Grid

OS Ordnance Survey

REST Representational State Transfer

SMIS Soil Management Information System

1. INTRODUCTION

5.10. Purpose

The following document describes the ‘use cases’ defined for the SMIS Analytics Toolkit software developed as part of the SMIS project (AHDB CP107D). The document’s primary purpose is to inform the end user about the defined uses of the system and the interaction flows required to achieve particular goals within its scope. On the developer side, the document will also serve as an implementation guide for the Analytics Toolkit, in particular for the design, development and installation of its front-end interface.

Additionally, as the inclusion of some use cases (see Section 0) will depend on validation of their viability during the remaining duration of the project, future versions of this document will effectively catalogue any additional use cases which are found to be viable and will therefore be included in the SMIS Analytics Toolkit.

5.11. Scope

This document describes the SMIS use cases without specifying implementation details beyond the overall flow of actor interactions the system is designed to allow. In particular, this means that both the back-end database, analytics and API functionalities are not discussed, and front-end interface details are not presented. Data gathering and manual curation activities undertaken within the scope of the SMIS project are also not discussed.

However, as the delineation of user roles (i.e. Operator and Administrator actors), as well as the classification of use cases used in this document result directly from certain system design and implementation decisions, a brief overview of the SMIS software design is provided in Section 2.

Additionally, a certain class of use case, the Established Query (described in Section 0) effectively consists of special cases of a more general use case, each intended to be used with its own custom interface and a more narrowly defined set of goals and user inputs. The list of Established Queries provided in the current version of this document is not intended to be exhaustive. These use cases and their associated interfaces will continue to be added to during the duration of the project as the viability (which is strongly dependent on gathered data) of various queries is verified and their usefulness validated. The document will be updated to include them as development continues.

6. OVERVIEW

The SMIS Analytics Toolkit software system can be considered to consist of several subsystems:

Parsing suite[3] – a set of command-line tools developed for the purpose of importing external datasets (e.g. Grower Data, Experimental / Research Project Data, Literature Data) into the SMIS database, capturing their internal relationships and ensuring their integrity. This component interfaces with the database back-end directly and is the primary route of interacting with the system for Administrator actors.

Database back-end[3] – a non-relational database responsible for storing the parsed data gathered within the scope of the SMIS project, as well as the results of analyses conducted by the Analytics back-end. This will be available to the front-end and Analytics back-end through a REST API.

Analytics back-end – a set of scientific computing tools developed for generating analyses and visualisations for the SMIS system. Other than relatively trivial analyses like summarising available data, this subsystem primarily focuses on machine learning algorithms designed to identify relationships within the gathered datasets or their subsets. This is available to the front-end through a REST API.

Web application front-end – a web application which provides the primary interface for the delivery of the main SMIS use cases. This accepts user input and accesses the Database back-end and Analytics back-end through a REST API in order to present (or generate) results of user queries. This is the primary route for interacting with the system for Operator actors, and as such it represents the main SMIS Analytics Toolkit user interface in general.

6.1. Actors

6.1.1. Operator

The 'Operator' actor is the primary user role defined for the SMIS Analytics Toolkit. The role represents users whose primary goal is the exploration and extraction of useful information from the SMIS database. This can consist of browsing and searching through the raw (and curated) datasets, visualising their contents, exploring "rule bases" which represent the information derived from the stored datasets using machine learning, constructing queries which can be used to generate more specific rule bases, or using pre-defined "Established Queries" to explore problems which have been identified as being both useful and possible to address, based on the data collected during the course of the SMIS project.

6.1.2. Administrator

[3] See the SMIS Database Architecture Technical Documentation for details.

The ‘Administrator’ actor is a user role responsible for populating the SMIS database. During the development stage, this role can therefore be fully identified as the developer. However, the data parsing suite is intended to be generic and well-documented, capable of being used to update and expand the SMIS database beyond the scope of initial development, if further data is acquired. At the same time, the wholly separate interface (command-line scripts on the SMIS server) and very different input and server access requirements clearly delineate the role of an Administrator as separate from a “normal” user or Operator.

It should be noted that the SMIS database is generally intended for periodic, rather than continuous, updates. This is because the addition of new data may invalidate the results of previous analyses, requiring them to be re-computed. This means that the Administrator actor is intended to be active relatively infrequently.

An additional point is that as the SMIS Analytics Toolkit has been designed with the assumption that it will be used internally within an organisation rather than made directly available to the public, certain functionalities which would normally be assumed to depend on an “Administrator” role (such as user account management or access control), are not part of the design and are therefore not covered by this document.

6.2. Types of use cases

Three types of use cases have been defined. The first two (Exploratory and Analytical use cases) overlap to an extent, due to being available through a common Web interface and used by the same actor.

6.2.1. Exploratory

Exploratory use cases are ones which involve the use of the SMIS Analytics Toolkit Web application to browse raw and curated datasets stored in the SMIS database. They do not lead to the generation or storage of new information and do not involve the use of the Analytics back-end.

6.2.2. Analytical

Analytical use cases are ones which involve the use of the SMIS Analytics Toolkit Web application to view summaries and visualisations of datasets stored in the SMIS database, as well as to create queries which result in the creation of novel rule bases. These use cases employ the Analytics back-end to generate summaries, visualisations and machine learning models and store this new data in the database. The primary purpose of storing the results is to avoid the need to re-compute results for repeated queries, allowing for easy exploration of previously generated rule bases.

6.2.3. Administrative

Administrative use cases are ones which involve the parsing of novel data, resulting in either the addition of whole new datasets, or appending new data to existing ones. These cases involve the use of the SMIS parsing suite by an Administrator.

7. USE CASES

7.1. Browse and export grower data

Primary Actor: Operator

Type: Exploratory, Analytical (extensions)

Description: The Operator browses the SMIS Grower datasets and filters them by an arbitrary selection of the available data fields. The contents of the database for a selected set of conditions can be inspected and visualised, allowing the Operator to assess the size and potential usefulness of the gathered data for a given set of conditions. The visualisations and raw data can be exported.

Main success scenario:

The Operator accesses the SMIS Analytics Toolkit application through a Web browser.

The Operator navigates to the Grower data view.

A paginated table containing grower data is displayed.

The Operator selects (for qualitative fields) or types search terms for each column of interest.

Database contents are filtered and displayed in a new paginated table.

The Operator inspects the filtered data.

(Optional) The Operator presses a button to export the filtered data as a spreadsheet and the generated visualisations (if any) as a PDF report.

Extensions:

4a. The Operator can hide extraneous or empty table columns for clarity.

6a. (**Analytical**) Timeline view.

6a1. The Operator selects the Timeline view option.

6a2. A paginated list of farm fields and associated scrollable timelines of filtered field operations and applications are displayed in a rotational context. Field operations are coloured depending on an assessment as to whether they were undertaken in or outside of LandIS generated workability days.

6b. (**Analytical**) Summary view.

6b1. The Operator selects the Summary view option.

6b2. A summary report describing the filtered data is displayed, including the dates for which data is available, soil types and crop hectarages (including diagrams showing the relative hectarages of various crops and varieties).

6c. By selecting a data row, the Operator may navigate to a **Rule Base view** (see Section 0.) filtered to display rule bases which make use of this data point.

7.2. Browse and export experimental data

Primary Actor: Operator

Type: Exploratory

Description: The Operator browses the SMIS Experimental datasets, selecting the specific Experimental datasets to be viewed and inspecting the collected data in a tabular form, with the option to filter it by an arbitrary list of conditions and, optionally, export any filtered data subset.

Main success scenario:

The Operator accesses the SMIS Analytics Toolkit application through a Web browser.

The Operator navigates to the Experimental data view.

A list of available Experimental datasets is displayed along with short summaries of the size and content of each.

The Operator selects an Experimental dataset to view.

A paginated table containing Experimental data is displayed.

The Operator selects (for qualitative fields) or types search terms for each column of interest.

Database contents are filtered and displayed in a new paginated table.

The Operator inspects the filtered data.

(Optional) The Operator presses a button to export the filtered data as a spreadsheet.

Extensions:

4a. Alternatively, the Operator may navigate to a **Rule Base view** (see section 0.) filtered to display rule bases which make use of this experimental data set.

5a. The Operator can hide extraneous or empty table columns for clarity.

9a. By selecting a data row, the Operator may navigate to a **Rule Base view** (see section 0.) filtered to display rule bases which make use of this data point.

7.3. Browse and export literature data

Primary Actor: Operator

Type: Exploratory

Description: The Operator browses the SMIS Literature datasets, viewing and inspecting the manually curated data summaries in a tabular form, with the option to filter them by an arbitrary list of conditions and, optionally, export any filtered data subset.

Main success scenario:

The Operator accesses the SMIS Analytics Toolkit application through a Web browser.

The Operator navigates to the Literature data view.

A paginated table containing manually curated summaries of Literature data is displayed.

The Operator selects/enters search terms for each column of interest.

Database contents are filtered and displayed in a new paginated table.

The Operator inspects the filtered data.

(Optional) The Operator presses a button to export the filtered data as a spreadsheet.

Extensions:

4a. The Operator can hide extraneous or empty table columns for clarity.

6a. By selecting a literature entry, the Operator may navigate to a **Rule Base view** (see section 0.) filtered to display rule bases which make use of this literature entry.

7.4. View rule bases

Primary Actor: Operator

Type: Exploratory

Description: The Operator can view the soil management and yield “rule bases” derived from the SMIS database using machine learning analyses. These rule bases can be viewed either in tabular form, as lists of issues vs. causes/solutions (for soil management rule bases) or yield vs. factors-affecting-yield (for yield rule bases) pairs along with their associated weights, or in graph form, rendering the same pairwise relationships in a graphical manner.

Main success scenario:

The Operator accesses the SMIS Analytics Toolkit application through a Web browser.

The Operator navigates to the Rule Base view.

The Operator chooses to view either Soil Management Rule Bases or Yield Rule Bases.

A paginated list of available Soil Management / Yield rule bases is displayed alongside queries used to generate them. Rule bases still in the process of being generated are greyed out and display a progress bar with an estimate of time to completion.

The Operator selects a rule base to view.

A graphical view of the rule base is displayed. This consists of a set of nodes and edges connecting them. The nodes represent soil management issues, causes and solutions (for soil management rule bases) or crop yield and factors affecting crop yield (for yield rule bases). Edges connecting the nodes represent relationships identified based on the SMIS database contents according to a specific query. The thickness of edges depends on the strength and the degree of confidence of the relationship (weight).

The Operator inspects specific relationships by clicking on edges which display the list of Experimental, Grower and Literature evidence for the relationship along with their associated weights in absolute and relative terms.

(Optional) The Operator presses a button to export the entire rule base (or a selected set of nodes/edges) in tabular form as a spreadsheet.

Extensions:

6a. Tabular view.

6a1. The Operator selects the Tabular view option.

6a2. A paginated table containing a list of issues vs. causes/solutions (for soil management rule bases) or yield vs. factor-affecting-yield (for yield rule bases) pairs along with their associated weights is displayed.

6a3. The Operator types in search terms for each column of interest in the table.

6a4. The rule base contents are filtered and displayed in a new paginated table.

7a. By selecting a piece of Experimental, Grower or Literature evidence from the list, the Operator may navigate to their respective database browse views (see sections 3.1, 0, and 0), filtered to display the selected entry.

7.5. Query rule base

Primary Actor: Operator

Type: Analytical

Description: The Operator constructs a query used to select data used by a machine learning algorithm to generate a novel, specific rule base.

Main success scenario:

The Operator accesses the SMIS Analytics Toolkit application through a Web browser.

The Operator navigates to the Rule Base view.

The Operator chooses to view either Soil Management Rule Bases or Yield Rule Bases.

The Operator constructs a query by selecting available categories (data fields from Grower and Experimental data and curated Literature keywords), optionally combining them with a manually entered search term. A query can contain an arbitrary number of such category/search term pairs.

The Operator selects the “Run Query” option.

Database entries are filtered according to the query and the Analytics back-end begins the process of generating a rule base based on the filtered data.

The Operator is redirected to the rule base view (see section 0), with the new query and its progress bar highlighted.

Extensions:

6a. If no data stored by SMIS matches the entered query the Operator is informed of this by a pop-up message and the query is aborted.

6b. If a rule base related to the same or equivalent query already exists in the database, the Operator is informed of this by a pop-up message and navigates directly to a view of that rule base (see section 0).

7.6. Established queries

Established Queries can be considered “special cases” of rule base queries as used in Section 0. They rely on the same database and Analytics capabilities, but instead of allowing for the construction of arbitrary queries, they have their particular input and output interfaces, each of which are tailored to only one very specific query, selected during the SMIS project based on their potential usefulness and the confirmed availability of necessary data. One example is the assessment of soil compaction risk.

Assess soil compaction risk: an example of an established query

Primary Actor: Operator

Type: Analytical

Description: The Operator assesses the soil compaction risk for a specific field (identified by NG code / OS Map Sheet) at a particular range of dates, optionally associated with a specific crop.

Main success scenario:

The Operator accesses the SMIS Analytics Toolkit application through a Web browser.

The Operator navigates to the Established Queries view.

The Operator selects the Assess Soil Compaction Risk query.

The Operator enters the NG Code and Map Sheet for the field of interest.

The low-resolution LandIS data stored by SMIS is accessed to provide the most likely soil type for the field. This is displayed in a drop-down menu and can be changed manually.

The Operator enters a start and end date for the Soil Compaction Risk assessment and (optionally) selects the crop of interest from a drop-down menu.

A colour-coded timeline of workability days based on soil type and weather data is displayed and inspected by the Operator.

Extensions:

7a. By selecting the assessment results, the Operator can navigate to the associated rule base (see Section 0) and view possible solutions to soil compaction, if any were identified.

7.7. Import grower data

Primary Actor: Administrator

Type: Administrative

Description: The Administrator imports standard Grower datasets (e.g. Gatekeeper XML / spreadsheet format) into the SMIS database.

Main success scenario:

The Administrator accesses the SMIS server through a command line interface and launches the Grower Data Import script.

The Administrator selects the option to import a new dataset and provides a path to a Grower data file.

The Grower data is imported on a row-by-row basis. Novel data fields or qualitative field values require confirmation, modification or rejection by the user (see the *SMIS Database Architecture Technical Documentation* for details).

A data import report is displayed and can be saved as a text file.

7.8. Extend grower data

Primary Actor: Administrator

Type: Administrative

Description: The Administrator imports tabular Grower data not stored in a standard format such as Gatekeeper XML. Such data needs to be manually associated with datasets already present in the SMIS datasets and serve to add extra fields not covered by Gatekeeper datasets imported earlier.

Main success scenario:

The Administrator accesses the SMIS server through a command line interface and launches the Grower Data Import script.

The Administrator selects the option to expand an existing dataset and provides a path to a Grower data file.

Data fields present among Grower datasets stored in the SMIS database, as well as those identified in the provided file, are listed.

The Administrator selects one or more pairs of data fields to be considered equivalents to be used for joining the new dataset with data present in the database.

The Grower data is imported on a row-by-row basis. Novel data fields or qualitative field values require confirmation, modification or rejection by the user (see the *SMIS Database Architecture Technical Documentation* for details).

A data import report is displayed and can be saved as a text file.

7.9. Import experimental data

Primary Actor: Administrator

Type: Administrative

Description: The Administrator imports novel Experimental datasets into the SMIS database using project-specific parsing modules grouped under a common interface.

Main success scenario:

The Administrator accesses the SMIS server through a command line interface and launches the Experimental Data Import script.

The Administrator provides a path to a data file.

The Experimental data is imported by a project-specific parser. Unexpected data fields or field values may require confirmation, modification or rejection by the user (see the *SMIS Database Architecture Technical Documentation* for details).

A data import report is displayed and can be saved as a text file.

3.10 Import literature data

Primary Actor: Administrator

Type: Administrative

Description: The Administrator imports spreadsheets containing manually curated Literature data gathered during the SMIS project.

Main success scenario:

The Administrator accesses the SMIS server through a command line interface and launches the Literature Data Import script.

The Administrator provides a path to a data file.

The Experimental data is imported on a row-by-row basis

A data import report is displayed and can be saved as a text file.

Appendix 11. Protocol for extracting data from Gate Keeper

Project overview

The aim of this 3-yr project (2015-2018) is to apply the principles of 'Big data' to provide best practice guidelines for sustainable soil management in horticulture. The Soil Management Information System (SMIS) will hold, manipulate and manage available sources of data/information pertaining to the specific effects of soil management practices on horticultural crop productivity and environmental protection (within a rotational context). The project will

- Use novel informatics techniques to create and then interrogate a rule base of soil management practices (and their outcomes) in different scenarios (soil type, crop, rotation, location, etc.)
- Provide users with a set of robust, evidence-based, best-practice soil management guidelines (and the likely consequences of applying them)
- Create an interactive platform, giving AHDB-Horticulture, and its growers, agronomists and land managers access to guidance on contextual, effective soil management practices

We already have whole farm (Gatekeeper) datasets from several growers, covering over 57,000 ha for the time span 2008-2015. This includes a number of horticultural crops [e.g. onions, celeriac, peas (vining and dried), French beans, spring beans] grown within a range of rotational contexts.

However, as highlighted by Martin Evans and John Chinn at the recent stakeholder workshop (June 2017), it is important that we bring in data from a larger number of field vegetable growers representing multiple grower associations.

Data requirement

We are looking for growers to provide us access to their Gatekeeper data. This will involve exporting specific data from your Gatekeeper data (process detailed in the Appendix) as an .xml file.

The specific data that we require pertains to soil management over multiple years (for a full rotational context) and should include: cropping history (including crop yields), soil analysis data, timing of all operations, products applied (and their timing), crop, variety, yield etc. We have found that the majority of the data that we require can be found under the following options in the Gatekeeper 'Analysis' menu list: 'Field Operations' (see Appendix, Step 5.2) and 'Previous Cropping' (see Appendix, Step 2.2). Under which the selection of specified headings (see Appendix, Step 2.4 and 5.4) are required to capture the precise data needed. The exceptions are where you hold some data e.g. field specific soil analyses or yields outside of the Gatekeeper system. In which case this data should be sent separately –either in hard copy format for later return, as scanned data or as an electronic copy. Your data will be held securely and anonymously within SMIS. No financial data is held.

Data output

In return your data (per field basis) will be linked to various outputs from LandIS (www.landis.org.uk), namely:

‘Machinery Workability Days’, the number of days when the land can be worked with acceptable risk of damage to soil structure.

‘Susceptibility to compaction’, the risk of soil compaction with each machinery/field operation based on the timing of operation relative to the machinery workability days.

‘Susceptibility to topsoil slaking’, a topsoil stability assessment indicating the risk of slaking (surface capping).

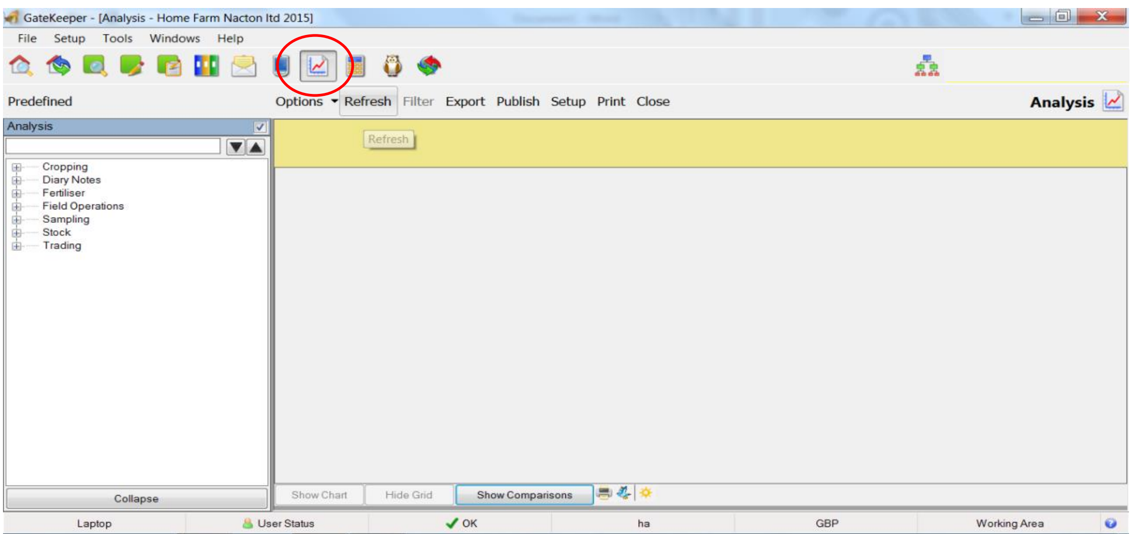
‘Potential for natural regeneration’, the potential for the soil to recover naturally from compaction and the time period (if any) over which this can be expected to occur.

Gatekeeper data extraction

Step 1: Refreshing Gatekeeper

Step 1.1:

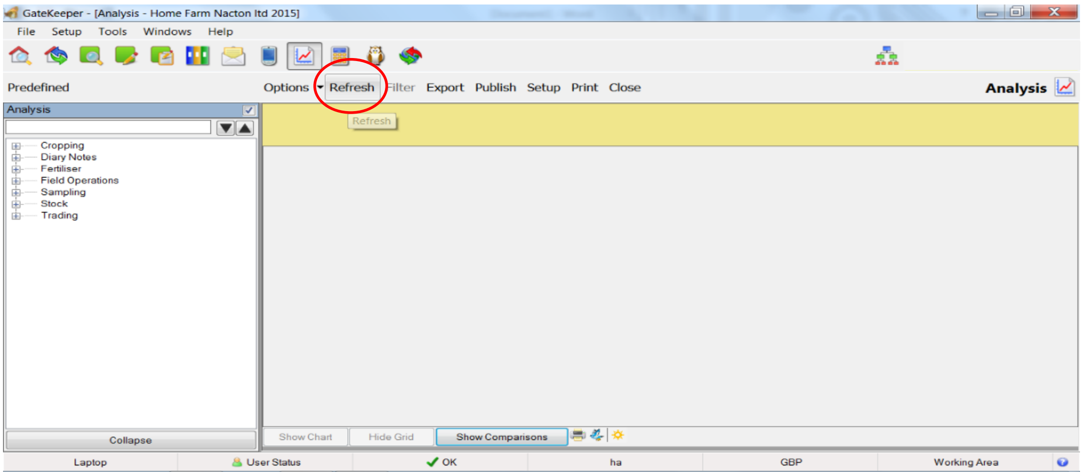
Click on the Analysis Icon (circled below). This will open up the ‘Analysis window’.



Analysis Icon

Step 1.2:

Click on the ‘Refresh tab’ (circled below). This opens the ‘Refresh Analysis Summary Data’ window (shown below). Select the years of interest that need to be ‘Refreshed’.

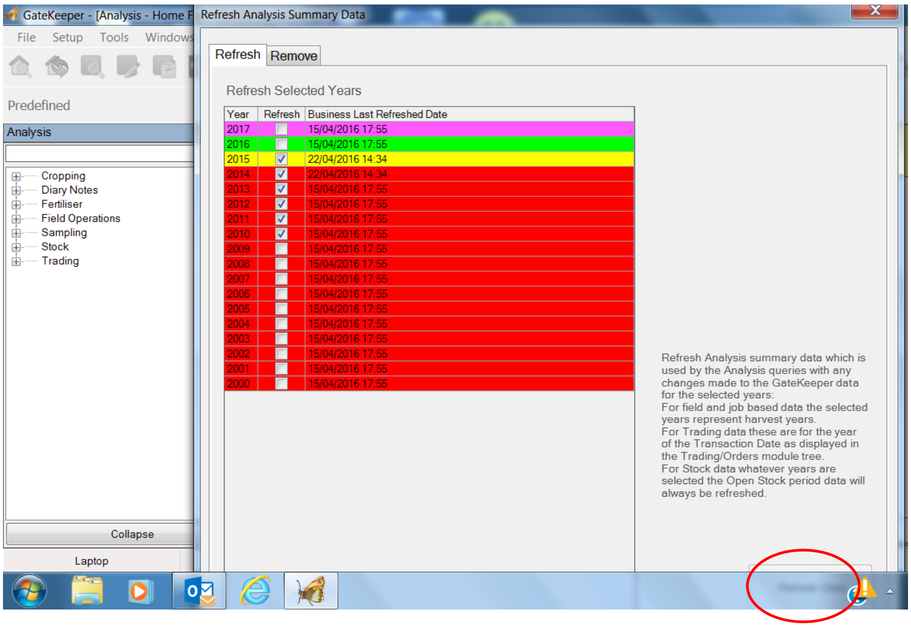


Refresh tab

Step 1.3:

Once the years have been selected, click the ‘Refresh button’ on the bottom right of the window (circled below) to refresh.

Note: If the ‘Refresh button’ is hidden by the task bar (as is the case below) press the TAB button on the keyboard to select the ‘Refresh button’ and then press ENTER. This will finalise the refresh action.



Refresh Analysis Summary Data

Step 1.4:

You can now close the 'Refresh Analysis Summary Data' window.

Step 2: Data preparation

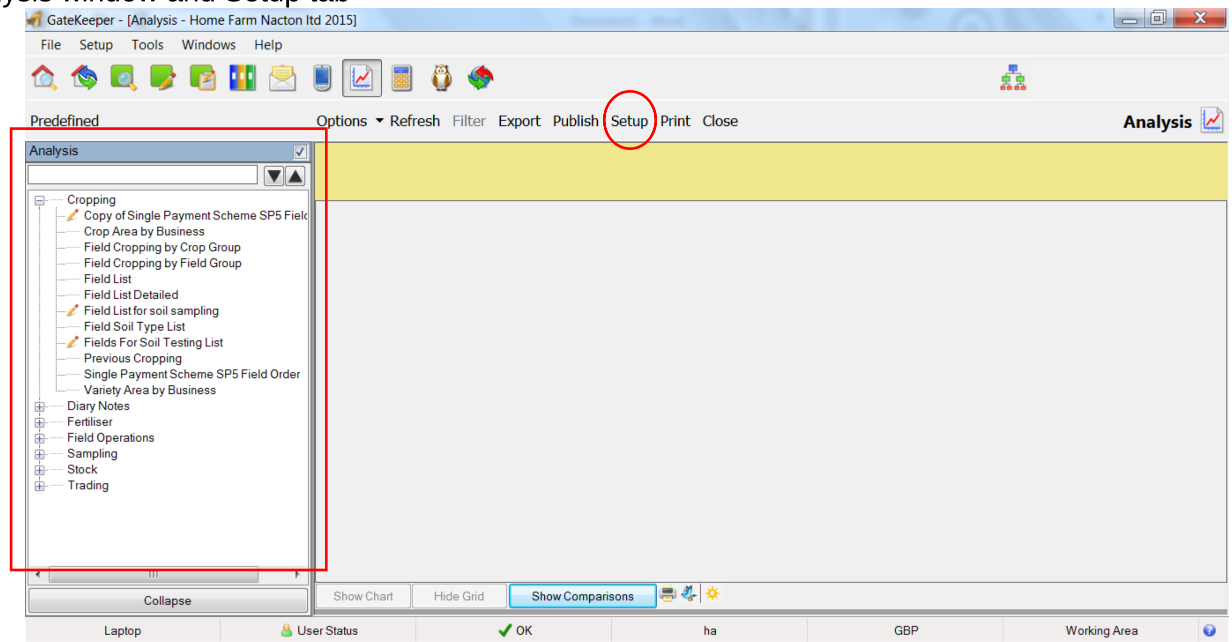
Step 2.1:

You have now returned to the 'Analysis window'. Under the 'Analysis menu list' on the left hand side of the 'Analysis window' (highlighted below) select the 'analysis' item that contains the data of interest.

Step 2.2:

Click the cross icon to expand the 'Cropping' item. Under this list you will find and select 'Previous Cropping'. You then follow Step 2.3 below.

Analysis window and Setup tab



Step 2.3:

As the original item cannot be edited a copy of the item must be made. With 'Previous Cropping' selected click the 'Setup tab' (circled above) and create a copy. Once a copy has been made, click the 'OK' button. You will now see this listed on the left in the 'Analysis menu list' as 'copy of Previous Cropping'.

Step 2.4:

With the 'Copy of Previous Cropping' selected, click the 'Options tab' [*Note: Do not click the drop down arrow*]. You then need to select the following list of items, excluding all other items:

Actual/Issued Date	Heading	Product Name

Application Area ha	Heading Category	Quantity
Crop	Heading Group	Rate per Application Area ha
Crop Residue	Heading Type	Split Number
Crop Sequence	Map Sheet	Status
Descriptor	NG Number	Total Yield
Field Defined Name	Official Area ha	Units
Field Group	OS Area	Variety
Field Number	Parent Field Name	Year
Field Reference	Part Field Reference	Yield Units

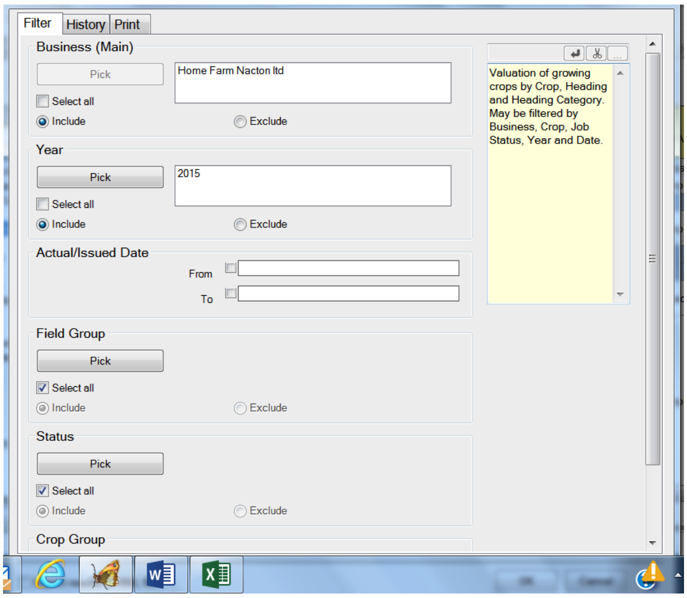
Step 3: Data extraction

Step 3.1:

Select the ‘Analysis menu list’ item for extraction. For example, select ‘Copy of Previous Cropping’. The pop up window below will appear.

3.2: Select the Years, Field Groups and Crops to be included.

[Note: To ensure that soil management can be evaluated in a full rotational context it is important to include as many years as possible. This also means that all crops are included such that a comprehensive range of rotational contexts can be captured. Depending on the processing capacity of the compute in use, it might be easier to extract the data in batches of only a few years at a time]



’Copy of Previous Cropping’ pop up window

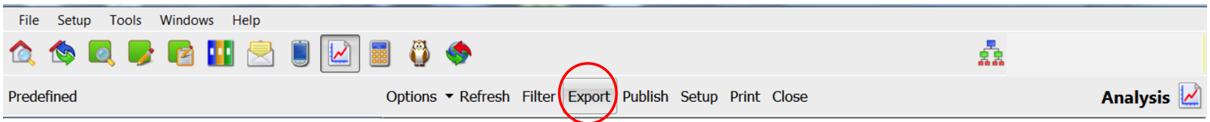
Step 3.3:

Once all the appropriate Years, Field Groups and Crops have been selected, click the ‘OK button’ at the bottom right corner. This will display the data available for export.

Note: If the ‘OK button’ is hidden by the task bar press the TAB button on the keyboard to select the ‘OK button’ and then press ENTER.

Step 3.4:

Click the ‘Export tab’ circled below and when prompted and save the file as an XML file.



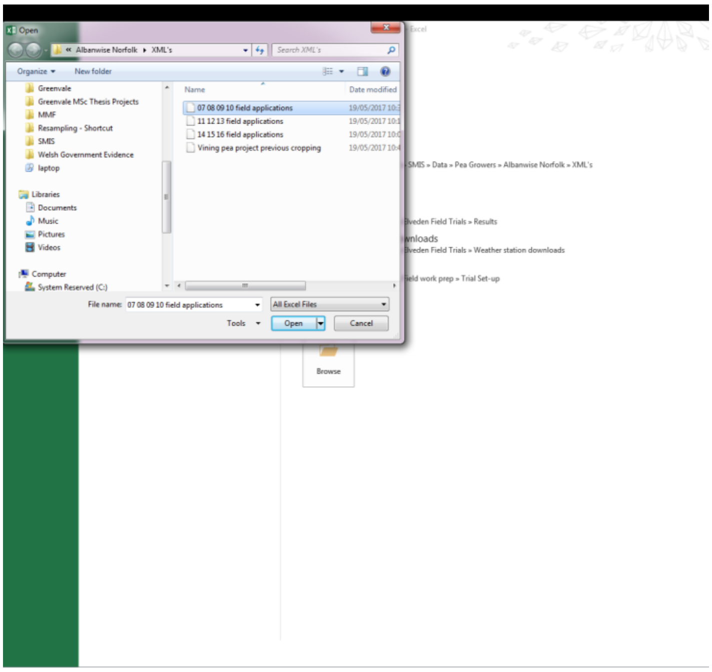
Export tab

Step 4: Checking the exported data

Step 4.1:

Once saved, open the XML file from inside EXCEL to check the output. To do this first open EXCEL and call the XML file from the EXCEL open file option (see below)

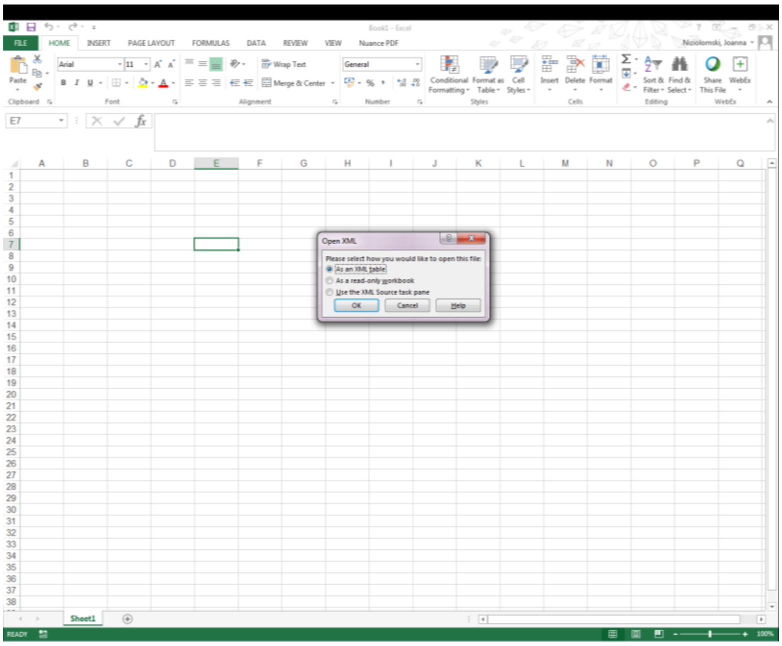
Note: Double clicking the XML file will not work.



EXCEL open file option

Step 4.2:

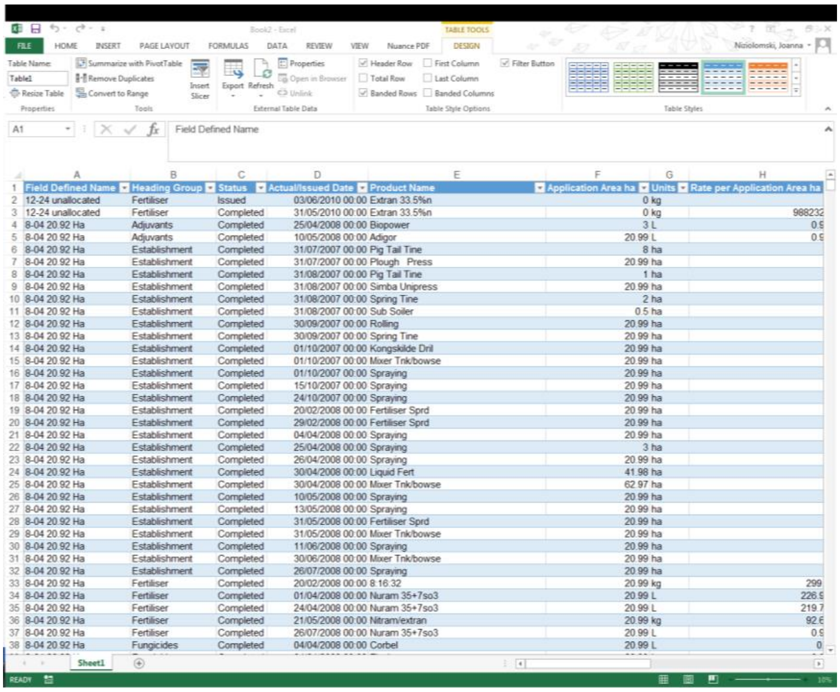
Select ‘Open as an XML file’ (see below).



Open as an XML file

Step 4.3:

Save the file as a normal EXCEL workbook format.



Step 5: Preparing the remaining data

Step 5.1:

Return to the Gatekeeper ‘Analysis window’. Under the ‘Analysis menu list’ on the left hand side (highlighted below) of the ‘Analysis window’ select the ‘analysis’ item that contains the data of interest.

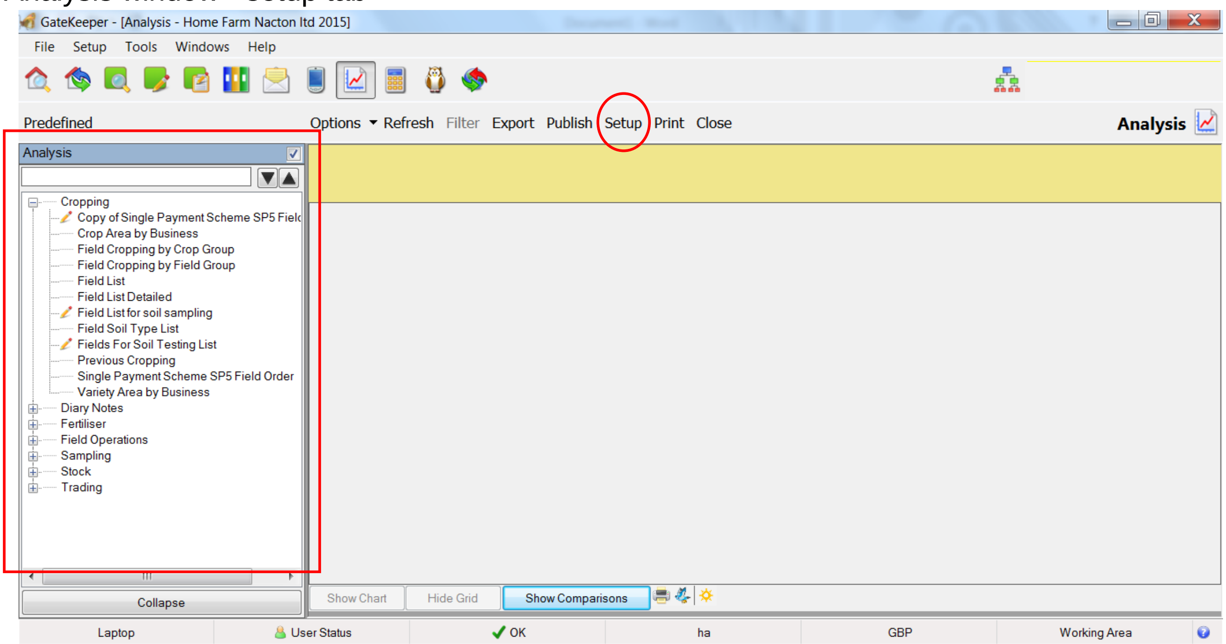
Step 5.2:

Click the cross icon to expand the ‘Field Operations’ item. Under this list you will find ‘Field Applications list’.

Step 5.3:

As the original item cannot be edited a copy of the item must be made. With ‘Field applications list’ selected click the ‘Setup tab’ (circled below) and create a copy. Once a copy has been made, click the ‘OK’ button. You will now see this listed on the left in the ‘Analysis menu list’ as ‘Copy of Field applications list’.

Analysis window - setup tab



Step 5.4:

With the ‘copy of Field applications list’ selected, click the ‘Options tab’ [*Note: Do not click the drop down arrow*]. You then need to select the following list of items, excluding all other items:

Actual/Issued Date	Heading	Product Name
Application Area ha	Heading Category	Quantity
Crop	Heading Group	Rate per Application Area ha
Crop Residue	Heading Type	Split Number
Crop Sequence	Map Sheet	Status
Descriptor	NG Number	Total Yield
Field Defined Name	Official Area ha	Units
Field Group	OS Area	Variety
Field Number	Parent Field Name	Year
Field Reference	Part Field Reference	Yield Units

Step 5.5:

You will now repeat Steps 3 to 4.

