

**Project title:** Development of a Soil Management Information System (SMIS)

**Project number:** CP107D

**Project leader:** R J Rickson, Cranfield University

**Report:** Final report (36 months)

**Previous report:** Annual report (24 months), December 2017

**Key staff:** Jane Rickson (Project Manager), Rob Simmons, Tomasz Kurowski, Fady Mohareb, Caroline Keay and Lynda Deeks

**Location of project:** Cranfield University

**Industry Representative:** Guy Thallon, FERA

**Date project commenced:** 01 November 2015

**Date project completed:** 31 January 2019

## **DISCLAIMER**

*While the Agriculture and Horticulture Development Board seeks to ensure that the information contained within this document is accurate at the time of printing, no warranty is given in respect thereof and, to the maximum extent permitted by law the Agriculture and Horticulture Development Board accepts no liability for loss, damage or injury howsoever caused (including that caused by negligence) or suffered directly or indirectly in relation to information and opinions contained in or omitted from this document.*

*© Agriculture and Horticulture Development Board 2019. No part of this publication may be reproduced in any material form (including by photocopy or storage in any medium by electronic mean) or any copy or adaptation stored, published or distributed (by physical, electronic or other means) without prior permission in writing of the Agriculture and Horticulture Development Board, other than by reproduction in an unmodified form for the sole purpose of use as an information resource when the Agriculture and Horticulture Development Board or AHDB Horticulture is clearly acknowledged as the source, or in accordance with the provisions of the Copyright, Designs and Patents Act 1988. All rights reserved.*

*All other trademarks, logos and brand names contained in this publication are the trademarks of their respective holders. No rights are granted without the prior written permission of the relevant owners.*


## 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 30 January 2019

[Name] R W Simmons

[Position] Senior Lecturer in Sustainable Soil Management

[Organisation] Cranfield University

Signature .....  ..... Date 30 January 2019

Report authorised by:

[Name] Professor A Thompson

[Position] Head of Cranfield Soil and Agrifood Institute

[Organisation] Cranfield University

Signature .....  ..... Date 31 January 2019

## Contents

Authentication .....	3
Contents .....	4
List of appendices .....	7
List of figures .....	7
List of tables.....	8
Grower summary .....	9
Headline.....	9
Background.....	9
Summary .....	10
Financial benefits .....	15
Action points .....	17
Science section.....	18
1. Introduction .....	18
1.1. Action points from Year 2 (2017) of the SMIS project .....	20
2. Materials and methods .....	21
2.1. Data collection and collation .....	23
2.1.1. Grower data .....	23
2.1.2. Literature review.....	23
2.1.3. Experimental data .....	24
2.2. Building the SMIS back end .....	25
2.2.1. A database technical documentation developed including the supported dataset types .....	25
2.2.2. Use Case scenario report describing data browsing and visualisation.....	25
2.2.3. Completed Web interface final report .....	25
2.3. SMIS user interface and analytics toolkit.....	25
2.3.1. Technical documentation and user manual report .....	25
2.3.2. Full access to the developed SMIS system granted to client for final feedback to be received .....	25
2.3.3. Final SMIS implementation and documentation to be formally handed over to AHDB .....	25
3. Results.....	26

3.1.	Data collection and collation .....	26
3.1.1.	Grower data .....	26
3.1.1.1.	Muddy Boots data.....	26
3.1.1.2.	Grower data and GDPR.....	25
3.1.1.3.	Integrating grower data with LandIS, RPA field boundaries and Met data .....	25
3.1.2.	Literature review.....	26
3.1.3.	Experimental data .....	30
3.2.	Building the SMIS back end .....	30
3.2.1.	A database technical documentation developed including the supported dataset types .....	30
3.2.2.	Use case scenario report describing data browsing and visualisation.....	31
3.2.3.	Completed web interface final report .....	31
3.3.	SMIS user interface and analytics toolkit.....	32
3.3.1.	Technical documentation and user manual report .....	32
3.3.2.	Full access to the developed SMIS system granted to client for final feedback to be received .....	33
3.3.3.	Final SMIS implementation and documentation to be formally handed over to AHDB .....	33
3.4.	Using SMIS on mobile devices.....	33
3.5.	Running SMIS.....	33
3.5.1.	Grower database.....	34
3.5.2.	Experimental database.....	34
3.5.3.	Literature database .....	35
3.5.4.	Rule bases .....	35
3.5.5.	Established queries .....	35
3.5.5.1.	Yield / increased productivity .....	36
3.5.5.2.	Compaction risk .....	36
3.5.5.3.	Foot rot index.....	36
3.5.5.4.	PCN level.....	36

3.5.5.5.	Cavity spot.....	37
3.5.5.6.	Soil erosion risk .....	37
3.5.5.7.	Other queries.....	38
3.5.6.	Case studies to demonstrate SMIS functionality and outputs .....	38
4.	Knowledge and technology transfer.....	39
4.1.	Project team workshops.....	39
4.2.	Stakeholder workshops.....	40
4.2.1.	Workshop June 2018.....	40
4.2.2.	Workshop July 2018.....	42
4.3.	Attendance at Grower Association and AHDB Panel meetings .....	44
4.3.1.	Year 3 (2018) .....	44
4.3.2.	SMIS project lifetime .....	44
4.4.	Compilation of KE events where SMIS was presented.....	45
4.5.	Articles for AHDB distribution to levy payers .....	45
5.	Discussion.....	46
5.1.	SMIS as an unprecedented source of data, information and knowledge on soil management in horticulture.....	46
5.2.	SMIS and the application of novel agri-informatics techniques.....	46
5.3.	Interrogating SMIS to better understand soil management issues.....	47
5.4.	The need for a multidisciplinary approach.....	48
5.5.	Engagement with industry.....	48
5.6.	Limitations of SMIS.....	49
5.7.	Final outputs from the project.....	50
6.	Conclusions.....	51
7.	Glossary.....	53
8.	References.....	56
9.	List of appendices .....	57

## List of appendices

Appendix 1. SMIS end user manual.....	57
Appendix 2. Protocol for extracting data from GateKeeper.....	57
Appendix 3. SMIS project Gantt chart summarising milestones and deliverables.....	57
Appendix 4. SMIS database architecture technical documentation .....	57
Appendix 5. SMIS system architecture.....	57
Appendix 6. SMIS use case documentation .....	57
Appendix 7. SMIS web interface report.....	57
Appendix 8. SMIS technical documentation and user manual report.....	57
Appendix 9. SMIS Steering Group .....	57
Appendix 10. Article for the Field Vegetable Review, July 2018 .....	57
Appendix 11. Presentation to the June 2018 stakeholder workshop.....	57
Appendix 12. Feedback from the SMIS workshop, June 2018 .....	57
Appendix 13. Presentation to the July 2018 stakeholder workshop.....	57
Appendix 14. Feedback from the SMIS stakeholder workshop, July 2018.....	57
Appendix 15. Events at which the SMIS project has been promoted.....	57

## List of figures

Figure 1. An example of SMIS output.....	12
Figure 2. Browsing the SMIS databases: an example showing the factors affecting yields of vining peas.....	13
Figure 3. SMIS: General system overview. ....	18
Figure 4. Soil management challenges in horticulture (after Rickson and Deeks, 2013).....	20
Figure 5. Scopus search alert output for identifying new relevant peer-reviewed literature.....	24
Figure 6. Classification of knowledge sources within the relevant literature (2019) .....	27
Figure 7. Number of literature items by country (2019).....	28
Figure 8. Distribution of relevant literature by AHDB Horticulture sector (2019).....	28
Figure 9. Distribution of knowledge items across each soil management challenge.....	29
Figure 10. Distribution of knowledge items by soil management solution (2019) .....	29
Figure 11. Opening page of SMIS .....	33
Figure 12. Participants intended uses of SMIS.....	41

## List of tables

Table 1. Soil management issues identified in CP107 (Rickson and Deeks, 2013) .....	10
Table 2. Increase in yields of crops grown in horticultural rotations due to improved soil and water management (2015 prices).....	16
Table 3. Action points from Year 2 Annual Report and corresponding activities in Year 3... ..	20
Table 4. Reporting of SMIS milestones and deliverables (with section numbers) .....	21
Table 5. Growers contributing data to the SMIS Grower database .....	26
Table 6. Extract from the literature review database, showing the number of relevant items and number of soil management challenges addressed in each source.....	26
Table 7 Soil management issues identified in CP107 (Rickson and Deeks, 2013) .....	36



## **Grower summary**

### **Headline**

The Soil Management Information System (SMIS) is an intuitive, relatively easy to use, web-based tool that can improve the productivity and competitiveness of UK horticulture. A comprehensive database on soil management problems and solutions has the potential to link grower data, experimental results and literature. SMIS outputs support data-driven decisions on sustainable soil management.

### **Background**

Soil management is at the heart of sustainable intensification as it has the potential to improve crop yield and crop 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 results from separate experimental trials 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 multiple 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 couldn'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 were unstructured, uncentralised and difficult to find and/or access. A real opportunity was identified 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 could hold, manipulate and manage such data in a way that could be interrogated to provide robust advice and guidance on the benefits of soil management practices, with regard to crop productivity and environmental protection.

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

Soil management issue	HDC Panel (2013)							
	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	✓		✓		✓			✓

## Summary

The Soil Management Information System (SMIS) is an intuitive, easy to use, web-based tool that brings commercial benefits to UK horticulture. SMIS can be accessed using Google Chrome at the following address: [www.smis.ahdb.org.uk](http://www.smis.ahdb.org.uk) (correct April 2019). A comprehensive database on soil management problems and their solutions, and functionality to link grower data, experimental results and literature have been developed. A user-friendly End User Manual (Appendix 1; separate document) has been written for advisors and growers to demonstrate how SMIS can be used to answer a wide range of queries related to soil management issues. The outputs from these queries can support data-driven decisions on sustainable soil management for yield optimisation, as demonstrated by examples in Figure 1 (how preceding crop affects Foot Rot Index) and Figure 2 (factors affecting the yields of vining peas).

The project has demonstrated that the principles of ‘big data’ can be applied to the diverse and dispersed sources of soil management data, knowledge and information in the UK horticultural sector. SMIS contains an unprecedented repository of horticultural grower data, based on Gatekeeper records (it is estimated over 40% of growers use Gatekeeper software). The database covers over 80 crop types, over several years within a cross rotational context, from a range of geographical locations across the UK. By including whole farm rotations, crops from other AHDB sectors are adventitiously included too such as cereals, oilseeds and potatoes. Currently there are 328,890 grower data records that can be interrogated within

SMIS. Novel agri-informatics approaches have been used to create, develop, operate and interrogate SMIS.



**Figure 1. An example of SMIS output**

A preceding crop of onions (green bar) has the most impact on foot rot index (FRI): potatoes have a negative effect (red bar) on the FRI value, indicating a reduction in foot rot risk.

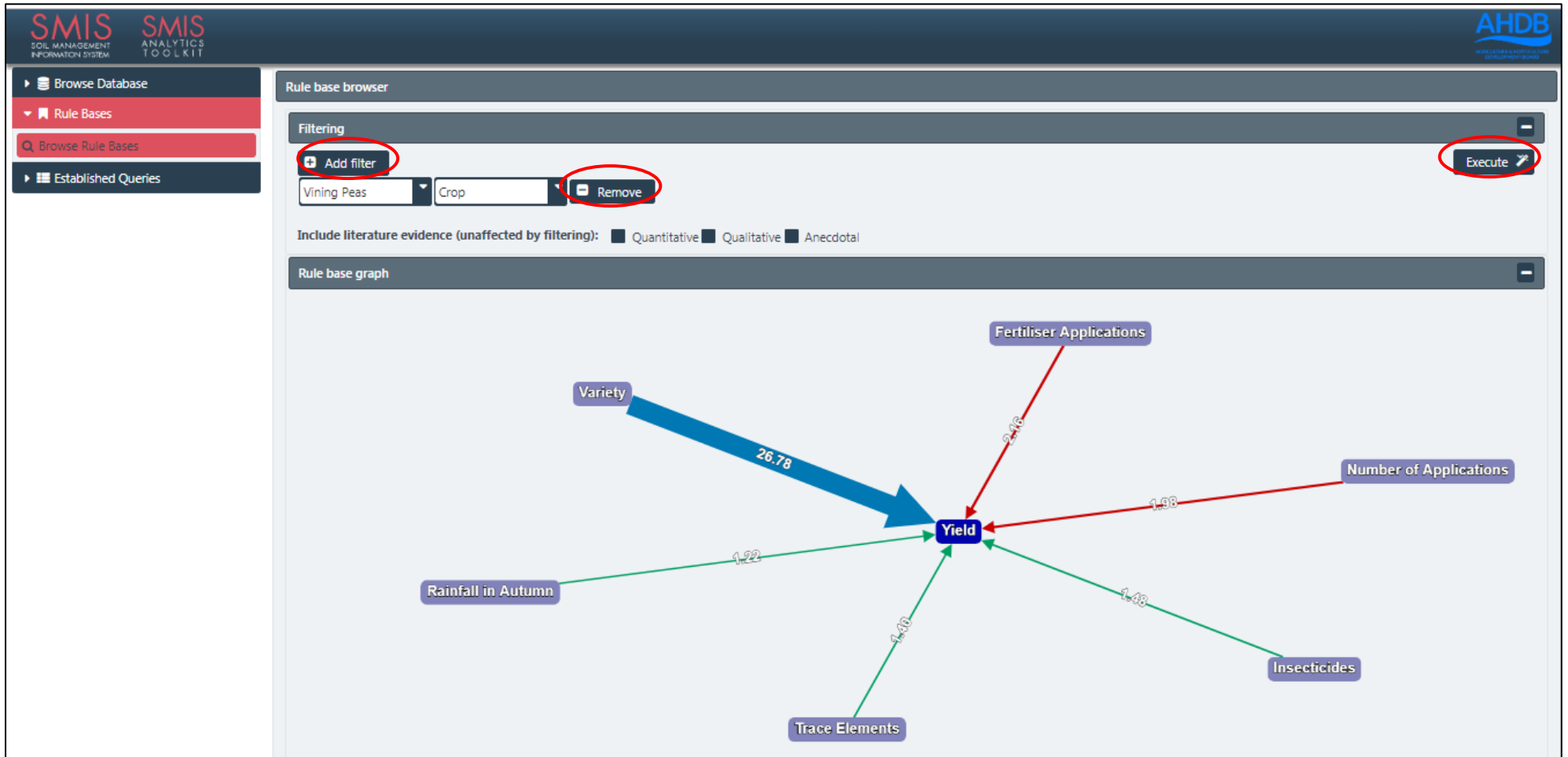


Figure 2. Browsing the SMIS databases: an example showing the factors affecting yields of vining peas

Where patterns in individual grower data are often obscured by site and time specificity, the pooled dataset can unearth relationships which were previously hidden. Patterns invisible in an individual data set are more likely to be revealed and can be used as the basis for best practice in sustainable soil management. These potential benefits can't be realised from data from one business on its own or even several businesses working together.

For the first time SMIS has allowed significant value to be added to this grower data by building functionality that has the potential to link this to other sources of knowledge where available, including experimental results and literature material. SMIS has improved the evidence base by incorporating previously diverse and dispersed sources of information and knowledge. SMIS is able to incorporate structured (e.g. Grower data), semi-structured (e.g. list of literature items, sorted by soil management issue) and unstructured (e.g. pdf documents within the Literature database) data sources.

The extensive, integrated, unique database can be interrogated by the end user to reveal causal relationships between soil management practices (e.g. operating outside the 'workability days' window) and outcomes (e.g. yields; compaction risk). The variety of the data included in SMIS means the number of queries and combinations of factors (crop/crop variety/previous crop/soil/year) that can be run is almost limitless. SMIS outputs are created by either browsing the Grower Data, Experimental Data and/or Literature Data held within SMIS, or by running an 'established query' on the data, such as factors affecting yield, foot rot index or PCN levels. SMIS outputs can then be analysed to identify the drivers behind soil management problems affecting yield optimisation and their solutions. The project has thus demonstrated that SMIS output can be used to inform on-farm decisions on horticultural soil management.

In terms of meaningful output from SMIS, strong, expected relationships between causes of soil management issues (e.g. use of fungicides in previous crop) and effects (e.g. foot rot index) are reassuringly found in SMIS. However, the 'big data' approach allows the end user "not to have any expectations, theories or hypotheses about the underlying relations, but rather use the observed patterns in the data to guide future decisions" (McAbee et al., 2017). Indeed, some of the observed patterns shown by querying SMIS are challenging to explain. These inferred relationships then lead to the question: "why?", prompting more investigations into the relationships shown.

As a specific example, one query shows that compaction risk in carrots is increased by the previous crop, notably leeks, winter wheat, savoy/green cabbage, sugar beet or spring barley. On the other hand, when white cabbage, potatoes and mustard seed precede carrots, compaction risk was reduced. This is a good example of where SMIS raises more questions

(i.e. why would growing savoy and green cabbage increase compaction risk in the following season, but white cabbage reduces it?).

Engagement with industry during a series of 'hands on', interactive Workshops gave SMIS positive reviews. The participants saw the value of SMIS in accessing large, complex and convoluted datasets to reveal causal relationships between multiple variables, including yield, so contributing to improved agricultural productivity. They valued having all the information in one place and the ability to see 'headlines' as well as drilling down to get to the finer detail. Using SMIS was considered to be relatively easy and intuitive, and produced outputs quickly and easily. Many of the suggestions on layout and operations have been incorporated into the current version of SMIS. The stakeholders particularly liked the ability to run different 'what if scenarios' as a basis to start discussions amongst growers. A number of applications of SMIS outputs were identified, including: analysing the impact of different surface tillage options; analysing the benefits of subsoiling; analysing options of increasing soil organic matter; analysing factors affecting yield; analysing options for reducing the risk of soil compaction and soil erosion. SMIS could also provide insight for challenges and validation of decision making – referred to as a "first stage triage" for growers in their decision making process. Looking ahead, the participants also commented that SMIS could be used to set future R&D priorities.

The SMIS project shows that 'big data' provides end users with much richer and more abundant information than previously available. Looking forward, the intrinsic system architecture and functionalities of SMIS allow it to be readily expanded to strengthen the statistical relationships found so far. These have to be significant before valid conclusions (and decisions) can be drawn. 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. The outputs of SMIS can be used to a) reaffirm current understanding of the effects of soil management practices and b) unearth new insights of possible causes of soil management issues and effects of soil management practices (requiring new research to validate those outputs).

The SMIS interactive platform provides AHDB Horticulture, and its growers, agronomists and land managers insights of contextual, effective soil management practices that can inform development of advice and guidance. Ultimately, the beneficiaries of SMIS will be farmers and growers.

## **Financial benefits**

The soil management advice and guidance given by 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 reseeded operations, subsoiling to alleviate compaction, releveling 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. Indicators of soil borne diseases (e.g. PCN counts, Foot Rot Index etc.) and how these are affected by field operations in a rotational context can also be investigated by SMIS. Managing these soil borne diseases will have direct benefits to growers.

The SMIS tool has been used to identify how field operations are linked to crop yields and soil degradation processes. This information will inform and justify future soil management decisions that maximise production (yield and crop quality) whilst protecting the terrestrial environment. By providing more effective advice and guidance on soil management, SMIS contributes to better soil health and system resilience. The benefits will accrue in terms of increased outputs (Table 2), and reduced inputs (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



## **Action points**

SMIS enables grower access to:

- interrogate the grower database, currently containing 328,890 items
- review the literature on soil management practices (and their effects) in horticulture
- run scenarios (including but not limited to the Established Queries) on a range of soil and crop related issues

## Science section

### 1. Introduction

The aim of the SMIS project is to provide comprehensive and coherent information to the horticultural sector in support of decisions on sustainable soil management. The intended outcome is to improve crop productivity, whilst reducing the causes and symptoms of environmental damage such as soil compaction. SMIS has been developed by applying the principles of 'big data' to the diverse and dispersed sources of soil management data, knowledge and information that are currently growing at an unprecedented rate. Patterns in individual grower data are often obscured by 'noise', which fades as the pooled dataset enlarges. Patterns invisible in an individual data set are more likely to be revealed and can be used as the basis for best practice in soil management. By improving the evidence base, SMIS will inform on-farm decisions on horticultural soil management

SMIS is an interactive platform that stores, represents and can potentially integrate, interrogate and analyse information from a range of sources in one place (a 'one stop shop' of information) (Figure 3). These include an unprecedented, anonymised database from horticultural growers (holding 328,890 items); the LandIS soil and environmental datasets; experimental results; and a wide range of literature material from academic and trade sources. The unique SMIS database covers a wide range of crops (often at variety level), different soil types and numerous field operations, over a number of years. SMIS operates over a seasonal and (more innovatively and uniquely) cross-rotational timeframe, allowing legacy effects of previous soil management decisions to be captured (these are often overlooked in limited time field trials and experiments).

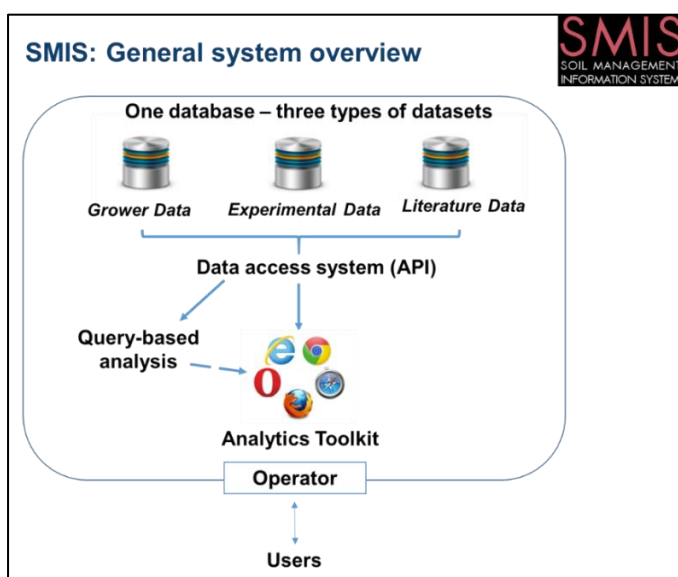
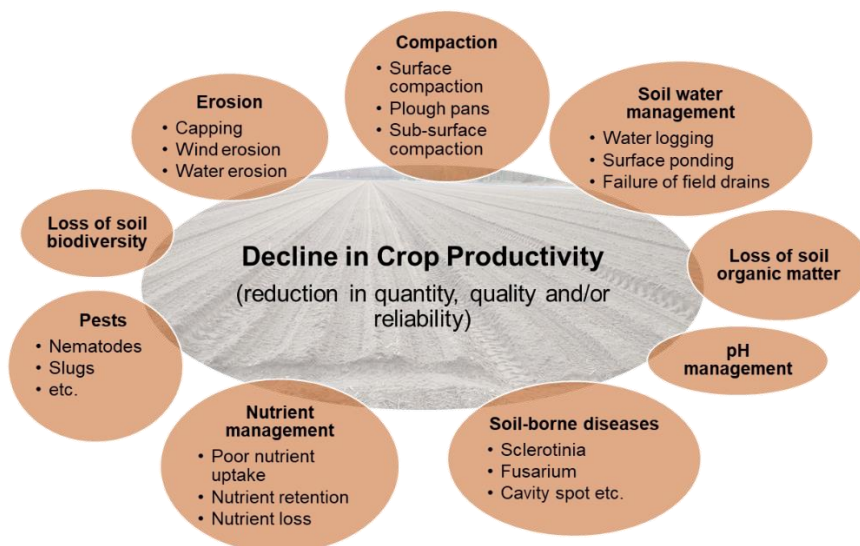


Figure 3. SMIS: General system overview.

New and emerging methods of data management and processing ('agri-informatics') allow meaningful interpretation of large datasets to unearth patterns undetected before. Novel informatics techniques are used to create and then extract patterns of 'cause and effect' regarding soil management practices (and their outcomes) in different scenarios (e.g. soil type, crop, previous crop, year, etc.). This is the SMIS 'Rule Base', which can be interrogated with specific queries related to soil management issues and challenges, as identified by a survey of growers in HDC Project CP107: "A gap analysis of soil management research and knowledge transfer in horticulture to inform future research programmes" (Rickson and Deeks, 2013; Figure 4).

The number of queries that can be run in SMIS is almost infinite, given the number of 'soil/crop/crop variety/previous crop/soil type/year' combinations in the dataset, but could include questions such as: What factors affect crop yield? What factors lead to soil compaction? What are the impacts of carrying out operations outside of 'workability' windows (i.e. when the soil is too wet)? How effective are fungicides on PCN levels? Examples of these queries can be found in the SMIS End User Manual (Appendix 1. ). SMIS is innovative in that it can unearth hidden yet valuable insights of the factors affecting soil management issues. For example, running the SMIS Established Query of 'Factors affecting yield" for celeriac reveals that 'previous crop' is the most significant factor affecting celeriac yield. By clicking on 'previous crop' further reveals that a previous crop of onions had the greatest impact on celeriac yield.

The SMIS interactive platform provides AHDB-Horticulture, and its growers, agronomists and land managers insights of contextual, effective soil management practices that can inform development of advice and guidance. It is estimated that better soil and water management can increase UK agricultural output by 5% or c. £500M/yr by 2020.



**Figure 4. Soil management challenges in horticulture (after Rickson and Deeks, 2013)**

**1.1. Action points from Year 2 (2017) of the SMIS project**

The following action points were identified in the SMIS Annual Report (2017). Year 3 (2018) activities have concentrated less on collecting the data and more on building SMIS, developing the analytics toolkit that can interrogate the database and then running SMIS. The key activities are listed in Table 3 and detailed in the relevant sections below.

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

Action Points from Year 2 Annual Report (2017)	Activities in Year 3 (2018)
1. Continue to source data / information / knowledge as input to SMIS database.	New literature and outputs from research projects have been inputted to SMIS as reported to Quarterly meetings
<ul style="list-style-type: none"> <li>• Grower data</li> </ul>	Growers continue to provide data and information to the SMIS data repository. A data pipeline of growers was established and members of the team used a standardised protocol to access the anonymised data from these growers (see section 2.1.1 below and Appendix 2)
<ul style="list-style-type: none"> <li>• Growers provide field records other than Gatekeeper (such as Muddy Boots).</li> </ul>	We have been unable to access Muddy Boots data (see 3.1.1.1 below)
2. Identify the queries to be run in SMIS. 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	See Appendix 1. SMIS end user manual

the SMIS database. These issues might include (and are not limited to):	
3. 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.	See Section on Rule Bases (see 3.5.4 below)
4. Develop the analytics toolkit to interrogate the database	See Section 2.3 below
5. Develop the SMIS User Interface (front end) – technical documentation will include the specification of the system and user manual	See Appendix 8. SMIS technical documentation and user manual report
6. 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.	See Section <b>Error! Reference source not found.</b> below
7. Consider technical and commercial implications of where SMIS will reside post project	See Section 5 Discussion and Section 7 Conclusions

## 2. Materials and methods

The structure of this section follows the activities in Year 3 of the project, as outlined in the Project's Milestones and Deliverables (Appendix 1. and Table 4.

**Table 4. Reporting of SMIS milestones and deliverables (with section numbers)**

	Milestone / Deliverable	Annual Report		
		2016	2017	2018
<b>1</b>	<b>Defining the scope of SMIS</b>	✓	✓	
	Analysis of growers' requirements of SMIS	✓	✓	
<b>2</b>	<b>Data collection and collation</b>			
	Literature review	✓	✓ (2.1.1; 3.1.1)	✓
	Journal papers	✓	✓	✓
	Grey literature	✓	✓	✓
	Research reports	✓	✓	✓
	Conference proceedings	✓	✓	✓
	Examples of data sources available from literature review		✓	✓

	AHDB Horticulture Soil Management Research Projects	✓	✓	✓
	CP107c Research Projects (ADAS)	✓	✓ (3.1.4)	✓
	CP107b Research Projects (Soil Association)	✓	✓ (3.1.2)	✓
	Previous AHDB Horticulture/HDC soil management research projects	✓	✓ (3.1.3)	✓
	Grower data	✓	✓ (3.1.4)	✓
	Grower data collection (e.g. Gatekeeper, Muddy Boots)	✓	✓ (3.1.4; 3.1.6)	✓
	Documentation outlining data required of growers/farmers (e.g. format of data and purpose of the data gathering)		✓ (Appendix 11)	(Appendix 2. Protocol for extracting data from GateKeeper)
	Integrating grower data with LandIS, RPA field boundaries and Met data	✓	✓ (2.1.3; 3.1.7)	See 3.1.1.3 below
	Expert opinion (Fuzzy Cognitive Mapping)		✓	
<b>3</b>	<b>Building SMIS backend</b>		✓ (2.2; 3.2)	
	A database technical documentation developed including the supported dataset types		✓ (3.2.1; Appendix 9)	Appendix 4. SMIS database architecture technical documentation
	Use Case scenario report describing data browsing and visualisation		✓ (3.2.2; Appendix 10)	Appendix 6. SMIS use case documentation
	Completed Web interface final report (including data access, and browsing)			Appendix 7. SMIS
<b>4</b>	<b>SMIS user interface and analytics toolkit</b>			
	Technical documentation and user manual report completed			Appendix 8. SMIS technical documentation and user manual report
	Full access to the developed SMIS system granted to client for final feedback to be received.			✓
	Final SMIS implementation and documentation to be formally handed over to AHDB.			✓
<b>5</b>	<b>PROJECT MANAGEMENT: Ensure the specifications, milestones and deliverables of the project are met, to time and to budget</b>			
	Contractual issues	✓		
	Revised contract	✓	✓ (3.4.1)	
	Development of a project risk register	✓		
	Revised risk register written and agreed		✓ (3.4.2)	
	Risk register reviewed	✓	✓	✓
	Quarterly updates to AHDB-Horticulture	✓	✓	✓
	Monthly updates to AHDB by telephone	✓	✓	✓
	Interim reports to AHDB-Horticulture	✓	✓	
	Final report to AHDB-Horticulture			✓
<b>6</b>	<b>Knowledge exchange activities</b>			

	Project Team Workshop (day event)	✓	✓	✓
	Project meetings/ Steering Committee meetings	✓	✓	✓
	Stakeholder Workshop	✓	✓	See Section 4.2 and Appendix 11; Appendix 12; Appendix 13; and Appendix 14
	Attendance at Grower Association Meetings	✓	✓	✓
	Compilation of KE events where SMIS is presented e.g. GA meetings / events	✓	✓	✓
	Articles for AHDB distribution to levy payers - eNews	✓	✓	Appendix 10

## 2.1. Data collection and collation

### 2.1.1. Grower data

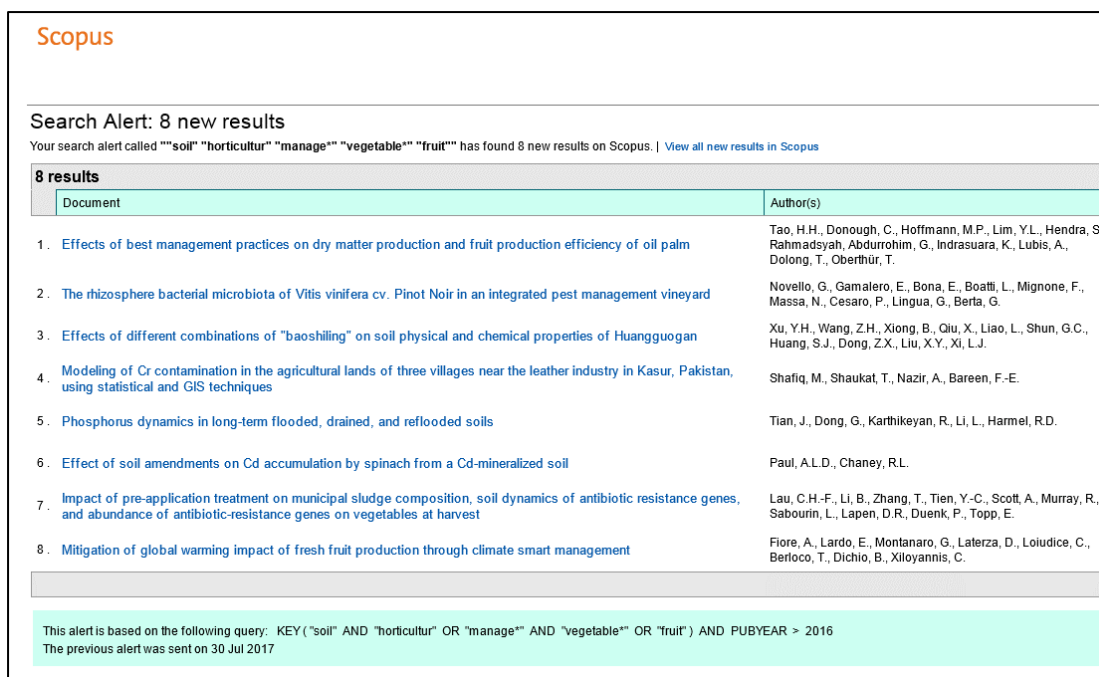
The methodology for extracting grower data to populate the SMIS database is described in the SMIS Annual Report (2017). Grower data in SMIS is mostly from Gatekeeper (used by approximately 40% of growers). Update on datasets received can be found in 3.1.1 below.

### 2.1.2. Literature review

The literature review report (SMIS Annual Report, 2017) describes the methodology for data collection from the literature. Sources of information covered in the review include:

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

The literature review has been regularly revisited for any available updates with new literature. A search alert was 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 5). New research projects were also investigated periodically for relevance to SMIS. This includes reviewing AHDB Horticulture, Defra and BBSRC research project webpages.



**Figure 5. Scopus search alert output for identifying new relevant peer-reviewed literature.**

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 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). The review was last updated in October 2018: all additional items from 2018 are included in the results of the literature review (see 3.1.2 below).

### 2.1.3. Experimental data

The data from a soil structure survey carried out in AHDB CP107C Precision Farming project has been incorporated into SMIS. Here, soil metrics were measured against field operations that have been undertaken on the fields sampled across their full rotational context. Ideally, sites will coincide with those captured in the grower database. Unfortunately, these were very limited in number. Even so, Paul Newell Price (ADAS) contacted two growers in the survey that had contributed Gatekeeper data to the grower database for their permission to



incorporate their soil structural survey results into SMIS. Unfortunately, research data from CP107C cannot be related to the Grower data: this is explained in section 3.1.3.

## **2.2. Building the SMIS back end**

The activities and outputs (deliverables) for this component of the SMIS development are reported in separate Appendices.

### **2.2.1. A database technical documentation developed including the supported dataset types**

See Appendix 4. SMIS database architecture technical documentation (Deliverable 3.1)

### **2.2.2. Use Case scenario report describing data browsing and visualisation**

See Appendix 6. SMIS use case documentation.

### **2.2.3. Completed Web interface final report**

See Appendix 7. SMIS .

## **2.3. SMIS user interface and analytics toolkit**

The activities and outputs (deliverables) for this component of the SMIS development are reported in separate Appendices.

### **2.3.1. Technical documentation and user manual report**

See Appendix 8. SMIS technical documentation and user manual report.

### **2.3.2. Full access to the developed SMIS system granted to client for final feedback to be received**

The SMIS system was made available to the client in September 2018, although the visualisation suite (Rules Bases) was delivered in October 2018.

### **2.3.3. Final SMIS implementation and documentation to be formally handed over to AHDB**

The final version and accompanying documentation was handed to the client in November 2018.

### 3. Results

#### 3.1. Data collection and collation

##### 3.1.1. Grower data

A spreadsheet 'SMIS Grower Datasets' was compiled from whole farm datasets provided by the following growers (Table 5). Project partners, PGRO supplied grower contacts as well as providing pulse / pea grower data. PGRO staff helped to explain and analyse the data in its raw form. Currently, the SMIS grower database comprises 328,890 individual entries.

**Table 5. Growers contributing data to the SMIS grower database**

Tompsett Growers Ltd	Sherwood Produce	Strawsons
Kettle Produce	Hardstaffs of Linby	James Foskitts
HUNTAPAC	Caley Farms Ltd	Jack Buck Farms
George Thompson Farms Ltd	Worth Farms Ltd	Hay Farming Ltd
Hammond Produce	Sam Rix	Albanwise
Parrish Farms	Stephen Barnes	

##### 3.1.1.1. Muddy Boots data

Despite requesting access to the raw data, MB refused access to their application programming interface (API), which is needed to upload Muddy Boots data into SMIS. They said access could be granted via the permission of individual growers who subscribe to MB. However, the only Muddy Boots data obtained was from Kettle Produce in the form of PDFs. This format is not suitable for integration with the parsing suite developed for SMIS. This is due to the difficulty of importing data stored as PDFs in general (i.e. importing entries cannot be done automatically) and due to there being no guarantees of the formatting not changing as it is designed to be human-readable, not for integration with software.

Muddy Boots (Paul Thomas, Senior Business Development Manager) suggested a manual data exchange file (e.g. Crop Walker) could be obtained from growers on Muddy Boots to upload data from MB to SMIS. However, this won't meet the need in the medium to long term of developing a sustainable integration plan with the Muddy Boots system. This is because Muddy Boots is unlikely to support future growers sharing their data via a consistent interface. In any case, Crop Walker is an old programme made by MuddyBoots (since replaced by Greenlight Grower Management, which is a cloud-based solution) which some users still use and which allows for exporting data (through data exchange files) to other tools like Gatekeeper.

It could be possible to integrate that data into the SMIS database, although it is arguable whether integrating data from an already obsolete (though still supported, for now) product

like Crop Walker would be supporting SMIS users and growers into the future. A free trial version of the new tool Greenlight Grower Management was tested to ascertain if it allows for exporting data to external software (such as SMIS). However, unfortunately, it just produces PDF reports and allows only limited data sharing through the web interface.

### **3.1.1.2. Grower data and GDPR**

With the introduction of GDPR Compliance in May 2018, it was necessary to check whether the gathering of data from growers and uploading it into SMIS was compliant. Cranfield University has a GDPR officer and Information Security Manager, who was consulted by the Project Manager. It was noted that the majority of grower data was collected before May 2018 when new GDPR rules came in. However, the Data Protection Act still applied at that time. To ensure data protection / GDPR compliance, a unique identifier was assigned to each data entry in the SMIS database as delivered to AHDB. Some data originates from a partnership / business, but all personal data from individual growers and experimental databases has been removed, so it is not possible to identify individual fields / farms / growers. Future data going into SMIS (post CP107D) would need explicit permission from growers to use their data and the ability to withdraw this at any time if they chose.

### **3.1.1.3. Integrating grower data with LandIS, RPA field boundaries and Met data**

LandIS data is now incorporated into SMIS in the form of 'look up' tables of soil and meteorological attributes. This is used to run the 'Rules Bases' and 'Established Queries' regarding soil compaction risk. The equivalent data for Scotland has been received from the James Hutton Institute and incorporated into SMIS. LandIS data is embedded in encrypted format for the current anonymised grower data within SMIS, as delivered to AHDB. Further details are in the 2017 Annual Report (November 2017) (page 50) and state:

"The 1km summary data 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."

LandIS 1km summary data has been integrated into SMIS and is provided for use within SMIS without additional charge.

### 3.1.2. Literature review

The 2018 review identified additional sources of literature, with the final review carried out in October 2018. In total, 86 different items pertinent to soil management in horticulture were found, with many referring to different crops, soil types, soil management challenges and solutions (Table 6). The following results show the final results of the literature review.

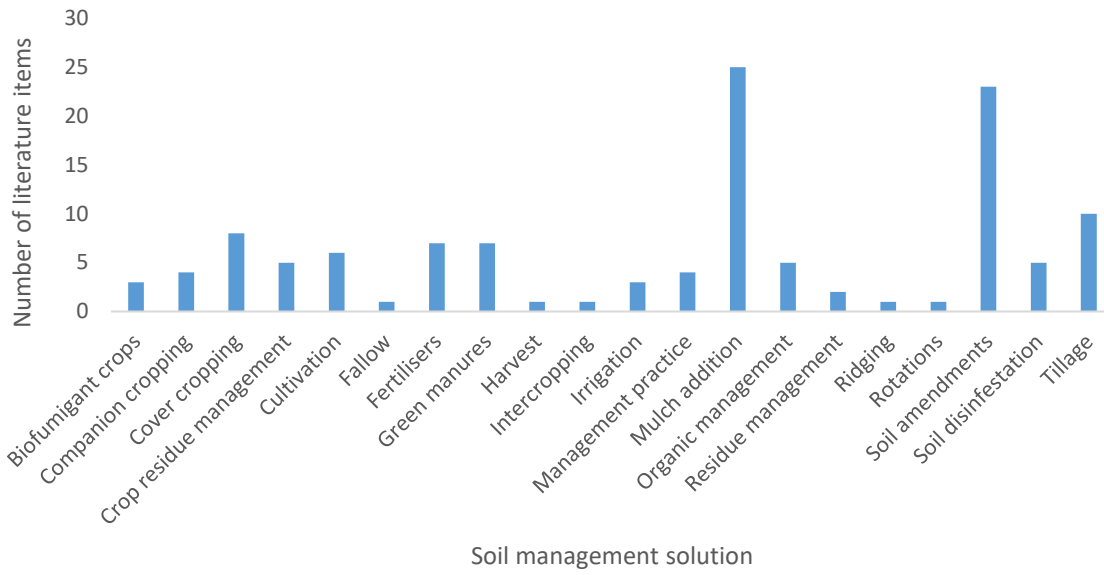
**Table 6. Extract from the literature review database, showing the number of relevant items and number of soil management challenges addressed in each source**

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		
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
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', <i>Acta Horticulturae</i> , 1076, pp. 143–150.	UK	Ornamental plants							x	x		x			
Arancon, N. Q. Edwards, C. A. and Lee, S. S. (2002) Management of plant parasitic nematode population by use of vermicomposts	USA	Tomato, bell pepper,		x							x		x		
Bailey, K. . and Lazarovits, G. (2003) 'Suppressing soil-borne diseases with residue management and organic amendments', <i>Soil and Tillage Research</i> , 72(2), pp. 169–180.	Global	Cross sector						x					x		
Barbara, J. (2010) Carrot cavity spot: (i) using quantitative PCR to predict disease in strawed crops; (ii) controlling moisture for optimum disease management	UK	Carrot											x		
Beniston, J. W., Lal, R. and Mercer, K. L. (2016) 'Assessing and Managing Soil Quality for Urban Agriculture in a Degraded Vacant Lot Soil', <i>Land Degradation &amp; Development</i> . John Wiley and Sons Ltd, 27(4), pp. 996–1006.	USA	Tomato, swiss chard		x						x					
Bešić, Z., Pantelić, M., Dabić, D., Iodić, S., Natic, M. and Iešić, Z. (2015) 'Effect of vineyard floor management on water regime, growth response, yield and fruit quality in Cabernet Sauvignon', <i>Scientia Horticulturae</i> , Elsevier, 197, pp. 650–656.	Serbia	Vineyards			x										
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', <i>Acta Horticulturae</i> , 1018, pp. 473–480.	Australia	Perennial fruit/veg							x	x					x
Biddlecombe, T. (2008) TF 179 — Pear: The effect of soil moisture on fruit storage quality	UK	Pear		x							x	x			

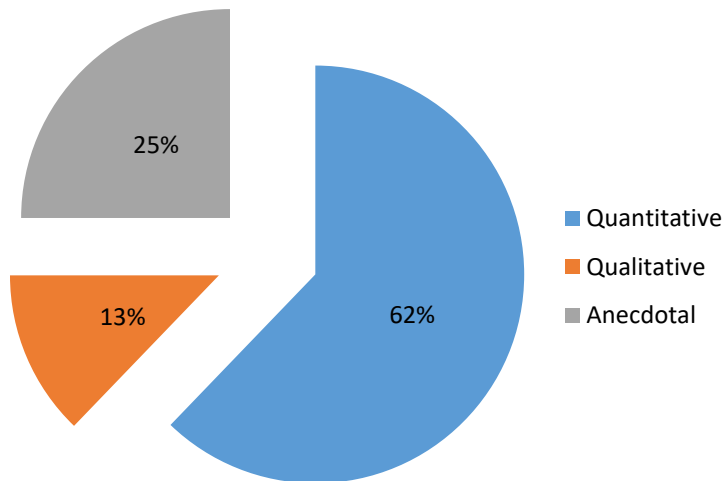
#### Knowledge classification key

Quantitative	Qualitative	Anecdotal
--------------	-------------	-----------

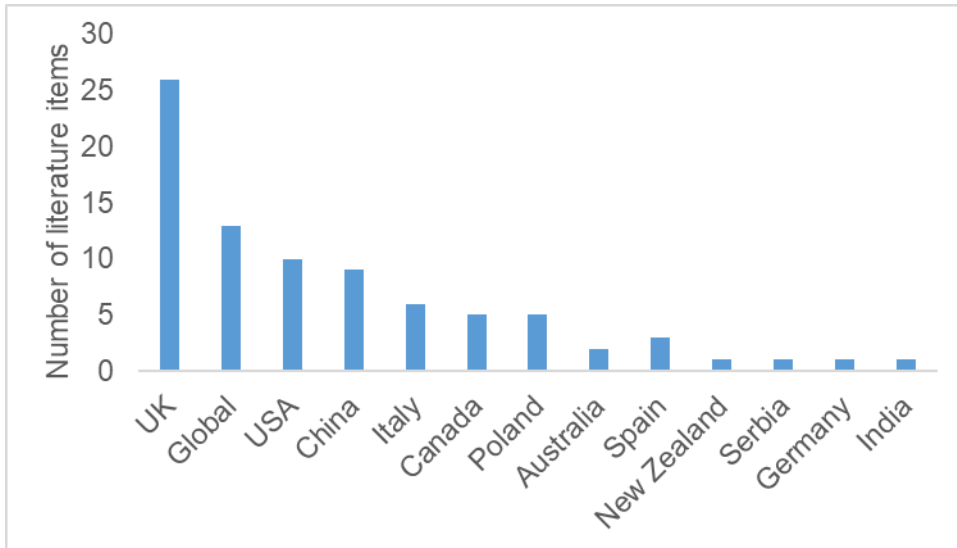
Quantitative references to soil management challenges were the most frequent, making up 62 % of the identified knowledge sources (Figure 6). The greatest number of knowledge items focus on research undertaken in the UK, with a good global distribution of other literature (Figure 7). Field vegetables, cross sector, tree fruit and protected edibles are particularly well represented (Figure 8). Regarding soil management challenges, the greatest amount of literature was found for soil-borne disease, followed by weeds and nutrient supply (Figure 9). The literature covers 20 broad soil management solutions with a focus towards generic 'management practice' (organic versus conventional) and rotation based solutions (



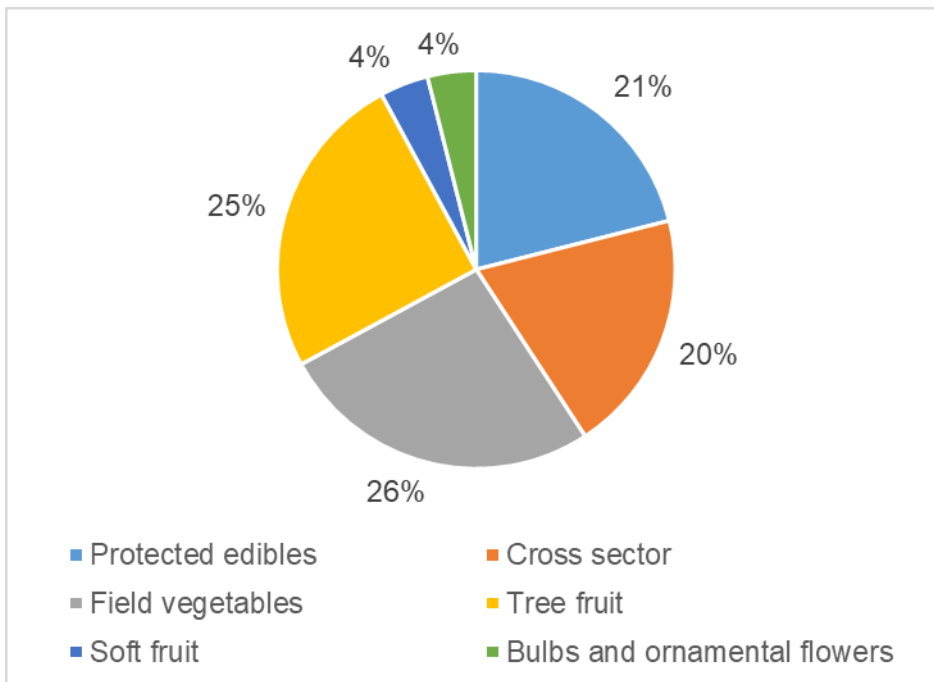
**Figure 10. All other solutions are evenly distributed across the identified literature**



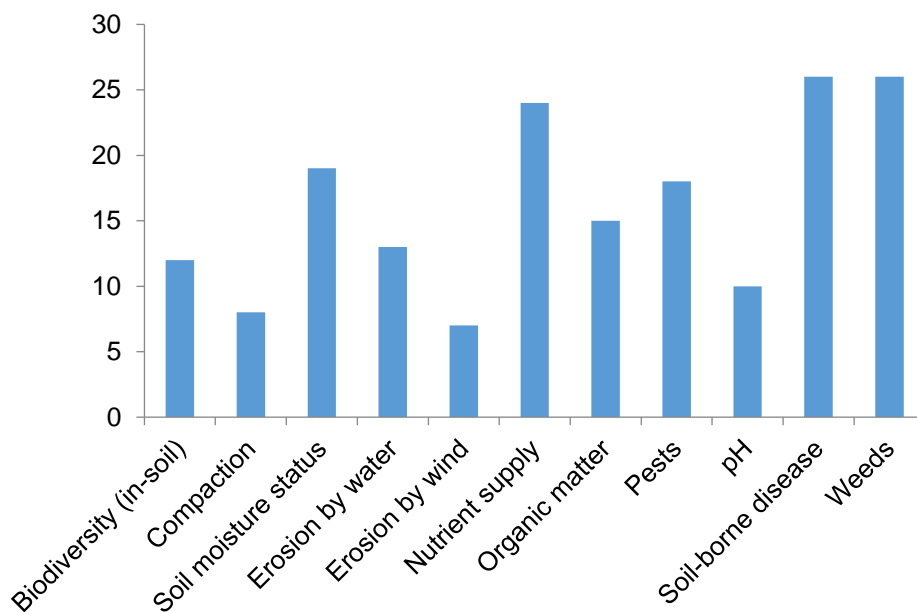
**Figure 6. Classification of knowledge sources within the relevant literature (2019)**



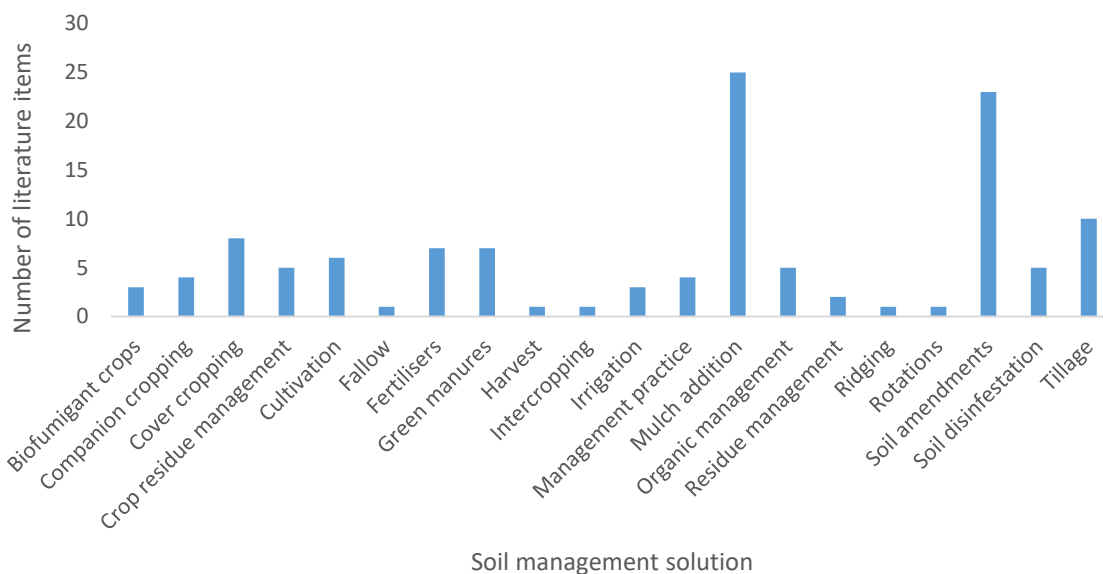
**Figure 7. Number of literature items by country (2019)**



**Figure 8. Distribution of relevant literature by AHDB Horticulture sector (2019).**



**Figure 9. Distribution of knowledge items across each soil management challenge.**



**Figure 10. Distribution of knowledge items by soil management solution (2019)**

The literature identified in the 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
- 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.

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 experimental data. This is discussed in section 3.5.4 below.

### **3.1.3. Experimental data**

There are currently 369 items of field experimental data in SMIS. The ADAS dataset from the soil structural survey within CP107C is now uploaded in SMIS and appears on the 'Experimental Data' tab in SMIS (see Appendix 1. SMIS end user manual). The soil structure and soil management survey was a single year survey (2016-17), so any year-on-year analysis is not possible. The experimental data along with other research data and the literature data integrates with the grower data using the visualisation suite / module, "Rule Bases" (see 3.5.4 below). Unfortunately, research data from CP107C cannot be related to the Grower data currently because:

- a) only 2 sites (fields) that appear in the ADAS trials also had Gatekeeper (Grower) data;
- b) there are no common (shared) outputs (e.g. yield) or variables that can predict/ explain the outputs (e.g. bulk density; organic matter) between the experimental and grower data; and
- c) the dataset is too small for any statistical relationships (e.g. linear regression modelling) to be run (i.e. it does not constitute 'big data').

## **3.2. Building the SMIS back end**

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

- a) SMIS database architecture technical documentation

See Appendix 4. This 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). The document provides an overview of the design and is intended to serve as an implementation guide for the developer. It contains an accurate description of the technical details of the system accessible to the end user. The document also explains the data upload steps and workflow.



The document has been updated from the version submitted in the Annual Report (2017).

b) SMIS system architecture

See Appendix 5. SMIS system architecture. The purpose of this document is to give an overall view of the SMIS software architecture, the rationale behind choosing the outlined programming environments, as well as presenting our vision for the main functionalities of the software framework.

**3.2.2. Use case scenario report describing data browsing and visualisation**

See Appendix 6. SMIS use case documentation. This 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.

**3.2.3. Completed web interface final report**

See Appendix 7. SMIS . This document describes the design of the interface of the Web application, referred to as the SMIS Analytics Toolkit (SMIS AT). This Web application is the platform for user interaction with data collected during the SMIS project, and includes features for data browsing, data analysis using machine learning methods, and the visualisation of both the analysis results and summaries of the collected data sets. This Web front-end depends on a database system and a representational state transfer application programming interface back-end (REST API), called the SMIS API, both described in separate documents (the Database Technical Documentation in Appendix 4 and the Technical documentation and User Manual report in Appendix 8), to deliver those functionalities.

The purpose of the document is to provide an overview of the front-end interface design, its structure and navigation options, without focusing on the technical details of the implementation.

### **3.3. SMIS user interface and analytics toolkit**

#### **3.3.1. Technical documentation and user manual report**

See Appendix 8. SMIS technical documentation and user manual report. This document describes the functionality and architecture of the software delivered as part of the Soil Management Information System (SMIS) project. Its purpose is to provide a summary of the overall system design, its intended means of deployment, descriptions of each of the system's major components alongside their individual architectures and dependencies, and an overview of how these components interface and interact with each other to provide the SMIS functionalities. Details on the means of system administration, configuration, and deployment are also included.

The document is intended to provide an accurate overview of the software as delivered and serves as a potential introductory document for an administrator or developer seeking to modify, expand, or re-configure SMIS software either at the front-end or at the back-end. Topics dealing with configuration and deployment are covered in separate sections to provide an effective manual for administrators/developers seeking to make simple changes within the scope of already implemented options without modifying the SMIS system software, which would require a deeper understanding of the design.

The document covers the software design of the individual components of the SMIS system and their interactions from a technical point of view, including a discussion of the technologies used (and the resulting requirements), the code organisation and implemented interfaces.

Overviews of the modules and classes that make up the software are included, but individual functions, methods, properties or other variables are outside of the scope of this document. Those lower-level elements of the implementation, of interest primarily to developers seeking to modify or expand the software, are documented via code comments including standardised tags which allow for automated generation of interactive, up-to-date, HTML-based documentation including hyperlinks, a form of documentation more conducive for software development. Scripts used for generating and viewing this documentation are an integral part of the SMIS software system to be delivered alongside it. Topics addressed in previous documents, in particular the Database Technical Documentation (Appendix 4), which covered the SMIS database design and generation process, and the SMIS Web Interface Report (Appendix 7), which covered the visual side of the interface views provided by the SMIS Analytics Toolkit, are covered with the focus limited to their interactions with other components and implementation details omitted from the previous documents.

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

This was delivered to AHDB in September 2018, but without the visualisation suite ('Rules Bases').

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

This was delivered to AHDB in November 2018 and included the visualisation suite ('Rules Bases').

## 3.4. Using SMIS on mobile devices

SMIS can work on a tablet (or even a mobile phone) via a web browser. However, because SMIS is intended for desktop web-browsers (and this is what we will perform testing and optimisation for), there are formatting issues (e.g. scaling of logos; sizing of graphs; layouts). Addressing these would require more time and effort than currently available on the project.

## 3.5. Running SMIS

Date	Crop	Variety	Field Operations	Product Name	Yield Units	Area	Quantity
2008-05-10	Winter Wheat	All Varieties	Adjuvants	Adigor			19.2
2007-07-31	Winter Wheat	All Varieties	Establishment	Plough Press			20.99
2007-07-31	Winter Wheat	All Varieties	Establishment	Pig Tail Time			8
2008-04-04	Winter Wheat	All Varieties	Establishment	Spraying			20.99
2007-08-31	Winter Wheat	All Varieties	Establishment	Spring Time			2
2008-04-26	Winter Wheat	All Varieties	Establishment	Spraying			20.99
2007-08-31	Winter Wheat	All Varieties	Establishment	Simba Unipress			20.99
2008-04-25	Winter Wheat	All Varieties	Adjuvants	Biopower			2.95
2008-05-13	Winter Wheat	All Varieties	Establishment	Spraying			20.99
2008-05-10	Winter Wheat	All Varieties	Establishment	Spraying			20.99

Figure 11. Opening page of SMIS

SMIS can be accessed at [www.smis.ahdb.org.uk](http://www.smis.ahdb.org.uk). The End User Manual (Appendix 1) describes the layout of SMIS and gives detailed instructions of how to run different queries within SMIS. The multifunctional structure of SMIS means it can be run in a number of ways by the end user. SMIS outputs are created by either browsing the grower data, experimental data and/or literature data held within SMIS, or by running an 'established query' on the data, such as factors affecting yield, foot rot index or PCN levels. These are described with illustrative examples in the End User Manual (Appendix 1), some of which were used in the End User Workshops held in June and July 2018. The end user can run queries on the grower database, the experimental data and/or the literature repository held within SMIS, either separately or in combination, when they are integrated in the 'Rule Bases' visualisation suite.

Here, the relationships extracted from the grower database are supported by a) the literature database and b) the experimental database within SMIS (where available) (see page 15 of Appendix 6. SMIS use case documentation).

Queries can be run to test expected outcomes (e.g. compaction risk increases in late harvested crops, due to wet soils in the autumn) or to evaluate the efficacy of different field operations (e.g. the effectiveness of growth regulators on crop yield). Uniquely, SMIS can also be used to find and display previously unseen, unexpected patterns in the data (so possibly prompting additional exploratory research). This unique and innovative aspect of SMIS is one of the most important outcomes of the SMIS project.

However, as mentioned, the outputs from SMIS are only as strong as the data held within the database(s) (and the relationships that can be drawn from the data). It should be noted that there are few direct links (e.g. shared outputs or predictor variables; corresponding locations) between the current experimental data and grower data. Also, the literature database deals solely with horticultural crops (as was the scope of the review). Since the grower database includes crops from whole farm rotations, many arable crops (e.g. cereals) are included, as well as horticultural crops. Finally, some soil management practices dominate the literature (e.g. mulching), but these are not captured in the grower database (i.e. Gatekeeper does not record the use of mulching as a soil management practice). Although the mechanisms/functionality for integration of all data have been created in SMIS (i.e. the visualisation suite and 'Rule Bases'), these reasons limit the degree to which the three datasets within SMIS are integrated currently.

### **3.5.1. Grower database**

The Grower database currently contains 328,890 individual items of interest, related to soil management in horticulture. The data originates primarily from Gatekeeper data supplied by farmers and growers. Queries that can be run on the Grower database and the interpretation of outputs are described in the End User Manual (Appendix 1).

### **3.5.2. Experimental database**

There are currently 369 items of field experimental data in SMIS, as shown in the End User Manual (Appendix 1). Some queries can be run independently on this dataset: for example, it is possible to run queries on the soil properties associated with a particular crop e.g. 'what is the mean level of Extractable P (mg/l) in cabbages on light soils' or 'which crops are more frequent on a particular soil type' etc. Experimental data can be connected to the grower database and literature database by the functionalities of the Rule Bases' visualisation suite

(see below). However, due to current limited availability of experimental data within SMIS, the linkages are not visually shown at present. This is because:

- a. Only two sites (fields) that appear in the Grower database have provided experimental data;
- b. There are no common (shared) outputs (e.g. yield) or variables predicting / explaining the outputs (e.g. bulk density; organic matter); and
- c. The dataset collected is too small for any statistical relationships (e.g. linear regression modelling) to be run.

### **3.5.3. Literature database**

There are currently 86 sources of literature on soil management issues and solutions. These can be searched to find information about particular crops, soil types, soil management issues and field practices. Examples are given in the End User Manual (Appendix 1).

### **3.5.4. Rule bases**

The Rule Bases of SMIS is a visualisation suite that aims to link the three data components of SMIS: the grower data, the literature and the experimental data. Further details are given in the End User Manual (Appendix 1).

### **3.5.5. Established queries**

The Established Queries function of SMIS aims to better understand and analyse the cause and effects of the horticultural industry's specific soil management challenges, as identified in the gap analysis of soil management research and knowledge transfer in horticulture by Rickson and Deeks (2013; Table 7). These issues can be addressed by SMIS, subject to the availability of appropriate and sufficient data.

**Table 7 Soil management issues identified in CP107 (Rickson and Deeks, 2013)**

Soil management issue	HDC Panel (2013)							
	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	✓		✓		✓			✓

### **3.5.5.1. Yield / increased productivity**

Gatekeeper records include crop yield which is a useful output indicator.

### **3.5.5.2. Compaction risk**

Gatekeeper does not record soil compaction directly (e.g. no measurements of bulk density are recorded in Gatekeeper). To overcome this, a proxy indicator of compaction was used, namely the use of subsoiling (which is recorded in Gatekeeper). In the absence of direct measurements, this remedial soil management practice was taken as reflecting 'soil compaction'. In Gatekeeper for example, 'establishment' refers to the cultivation method used – this can indicate alleviation of previous season compaction issues (e.g. use of 'subsoiling'). Equally, use of machinery outside of machinery workdays was also taken to indicate compaction risk.

### **3.5.5.3. Foot rot index**

Incidence of foot rot is taken from the recordings of foot rot index from pea grower data. Although only limited data has been provided by growers, an established query has been built and can run in SMIS. An example is shown in the End User Manual (Appendix 1).

### **3.5.5.4. PCN level**

SMIS includes some information on the counts of PCN in potatoes. This established query has now been built and run (See Appendix 1), but it should be noted that there are very few data points (based on only 2 grower data sets)

#### **3.5.5.5. Cavity spot**

No direct data on the incidence and severity of cavity spot was available. This is because the relevant data is linked to pack house data management systems rather than held within Gatekeeper. However, this pack house data was not provided by growers even though requests were made on a number of occasions. SMIS is set up in such a way that such data could be incorporated in a subsequent iteration.

In addition, raw experimental data was obtained from AHDB project FV373. Via the British Carrot Growers Association (BCGA) an email was circulated to those growers who had participated in FV373. The intention was to obtain Gatekeeper data (and hence the rotation context) for the fields sampled during FV373. However, no growers came forward.

Consequently, we were required to investigate the use of data proxies for cavity spot incidence (e.g. when carrots are harvested early in wet months), but no relationships could be found. It was concluded that there was insufficient data in the system to generate statistically significant relationships on cavity spot incidence, which limited the ability to run queries on the factors influencing it. However, SMIS is set up in such a way that if the grower field and packhouse data should become available that meaningful relationships could be identified.

#### **3.5.5.6. Soil erosion risk**

Measurements of soil erosion or soil erosion risk are not made in Gatekeeper (grower database). The SMIS database was therefore analysed to ascertain whether appropriate metrics are available in the database to run a query on soil erosion risk. Assuming the focus is on the most common form of soil erosion – soil loss by water (rather than by wind, co-extraction on root crops or farm vehicles, or by tillage) soil erosion risk can be estimated from:

- Rainfall erosivity (intensity and duration). SMIS contains information on annual precipitation data from the Met Office, but this is poorly correlated to erosion events;
- Soil type (at least according to Defra / RB209 classification at the coarsest resolution that includes sands, silts and clays). This will indicate the erodibility of the soil to erosion processes;
- Slope steepness and length. These are missing metrics from the Grower database and are not recorded by the majority of growers;
- Cropping pattern. This information is available from some of the grower data, in terms of crop type, previous crop and timing of operations.
- Presence of existing soil erosion control measures (e.g. grass buffer strips, retention of residues as mulch). This is a missing metric and is difficult to infer from other metrics (e.g.

as has been done for compaction, using the use of subsoiling as a proxy metric for compaction occurrence).

After due consideration, it was accepted that the data within SMIS is inadequate to run a query that identifies the likely factors affecting soil erosion / soil erosion risk.

#### **3.5.5.7. Other queries**

It was hoped that SMIS could be used to find the factors affecting other outcomes, such as yield/crop quality; yield gap and yield reliability. However, these are missing metrics within the SMIS database.

#### **3.5.6. Case studies to demonstrate SMIS functionality and outputs**

See End User Manual (Appendix 1) that includes a number of illustrative case studies.



## **4. 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.

### **4.1. Project team workshops**

A demonstration of the SMIS prototype was held at Cranfield University on 27<sup>th</sup> March 2018. Staff from AHDB, Martin Evans (Chair of AHDB's Field Vegetable Panel) and the Cranfield project team were present. The purpose of the meeting was to demonstrate the functionality of the prototype SMIS (the full demonstration was not scheduled until the end of May 2018). A subset of the database of 6-8 growers was used for the demonstration, representing 260,951 data entries and over 6,000 hectares. Other data collected by the team was waiting to be uploaded into SMIS.

The feedback received from the workshop was positive. The interactive aspect of SMIS was very welcomed and using the programme was intuitive. The system was seen to encourage curiosity amongst growers. The 'big data' concept and analytics may provoke new ideas for soil management. The anonymity of the data was raised as a concern, but reassurance was given that all entries from the grower database had been anonymised and no individual, company or location could be identified.

The platform running SMIS was discussed to see if it could be accessed via iPhone / iPad or smart phone. This is possible (as SMIS is a web based application), but the width of the spreadsheets displaying the data does not lend itself to viewing on a small screen.

Feedback included the need to 'tidy up' the data with regard to labelling and descriptions. It is important that the terminology is consistent, especially regarding crops and management options. One example was to reconcile items labelled as 'vining peas', as opposed to 'peas: vining'. It was agreed that the vocabulary will become easier to recognise with more input (and the SMIS dictionary expands). Parsing the different data sources is automatic, until a new, inconsistent or unrecognised entry (i.e. anomaly) appears. Then parsing has to be done manually until that entry is common and readily recognised.

It was suggested that user guidance could be provided via a webinar/ webcast rather than text document. This was acknowledged, but an illustrative user manual (with many screen shots of actual queries run in SMIS) has been produced (End User Manual; Appendix 1).

The next steps are to discuss the development / design of the User Interface to ensure it is compatible with AHDB internal systems.

## 4.2. Stakeholder workshops

### 4.2.1. Workshop June 2018

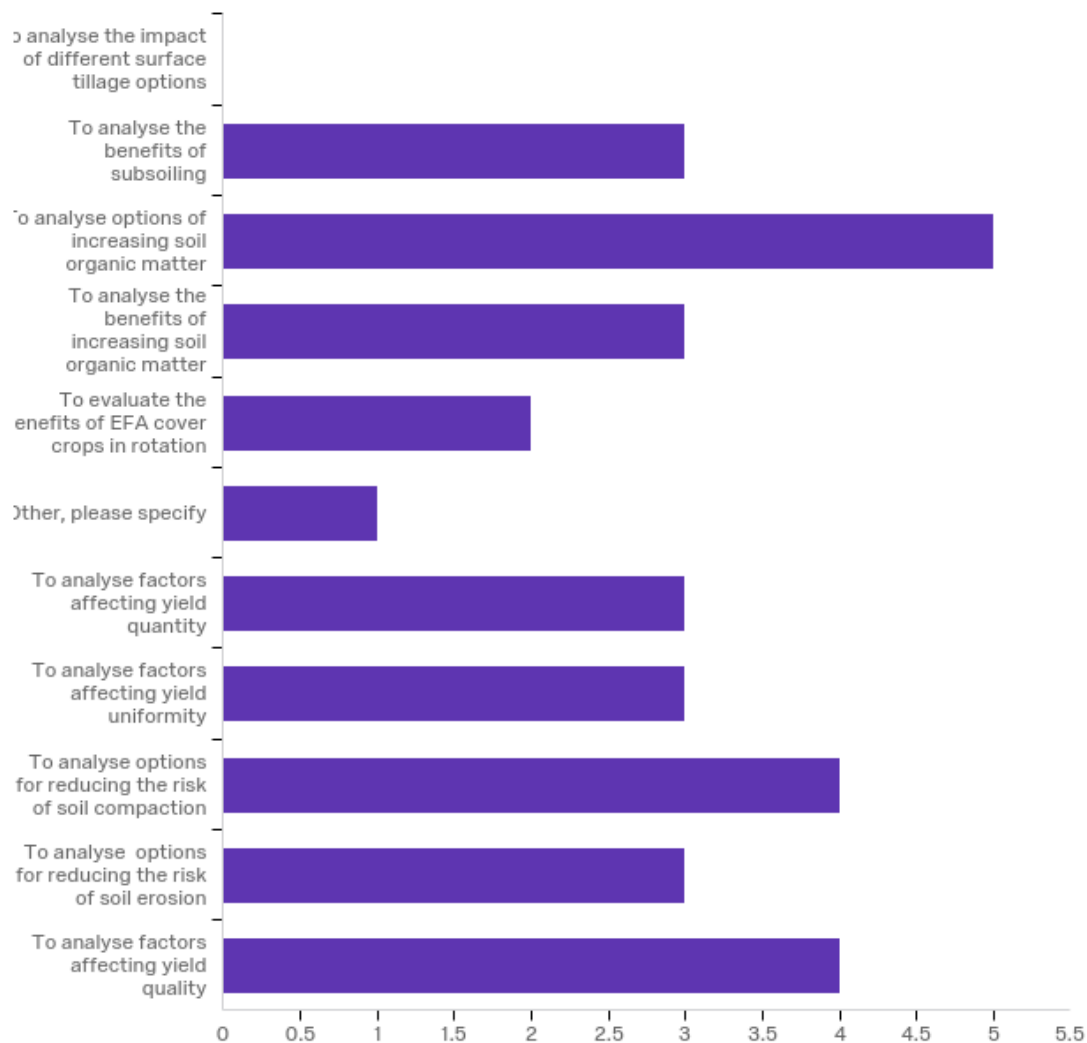
A technical workshop was arranged for AHDB staff at Cranfield University on 08 June 2018. The purpose was to introduce the project and run a hands on practical of how SMIS can be used to improve soil management advice and guidance. Presentations for the workshop are shown in Appendix 11. Presentation to the June 2018 stakeholder . The prototype SMIS was run for specific soil management challenges, queries etc. The feedback received on using SMIS and its outputs would be used to develop and improve the system further. This feedback was recorded and responses to it are described in Appendix 12. Feedback from the SMIS workshop, June 2018.

Many of the comments referred to the ways of navigating through the system, or the aesthetics of the displays. It was possible to modify SMIS to address these issues (e.g. select a range of dates of interest; putting the drop down menus of crops or crop varieties in alphabetical order; ensuring terminology is consistent (e.g. does 'bulbs' refer to onions or narcissi?); separating organic from inorganic production; In many cases, the terms used are dictated by the categories / headers in the original source, Gatekeeper datasets. For example, one header is simply "applications" which could refer to applications of herbicide, insecticide or fertiliser etc. Without talking directly to every grower (which was possible with some growers, but limited by the amount of time they could afford) to decipher what is meant by 'application' in every data row, inevitably some overlap / different meaning will occur. The definitions of these terms will inevitably vary from grower to grower. It is recognised that some operations may be included twice (i.e. double counting) – e.g. 'insecticides' may also mean (and appear as) 'applications'. One user suggested to add error bars to (mean) yields where available. However, 'mean yield' would be meaningless given the range of conditions where each crop has been grown. However, benchmarking for a given soil type and geographical location has potential for future development of SMIS.

After the workshop, participants were asked to complete an on-line Qualtrics survey to record their impressions of using SMIS. Overall, the participants were very positive about using SMIS with 4/5 rating their experience and level of satisfaction in using SMIS as 'OK' (5 options available: 'Very good'; 'Good'; 'OK'; 'Poor'; 'Very poor'). (Given this was using the prototype version, still with 4 months development to be completed, this was encouraging to the team). All participants rated the speed of application as being Good or Very Good. With regard to the output of SMIS, the participants were satisfied that it answered the question they had asked. The participants also had the opportunity for specific comments: these are recorded in Appendix 12. All participants said they felt confident using SMIS following the workshop –

demonstrating the system is intuitive and easy to use. They also listed a number of applications where they would use SMIS (Figure 12).

The End User Manual (Appendix 1) incorporates many of the comments to improve the end user experience. A suggestion was made to make a YouTube video for end users: a good idea, but this is not an agreed deliverable. The End User Manual includes a number of illustrations / demonstrations of SMIS applications.



**Figure 12. Participants intended uses of SMIS**

#### **4.2.2. Workshop July 2018**

The SMIS Demonstration Workshop was held at Cranfield University on 10 July 2018. This is a very busy time of year for many growers, so the attendance was disappointing (despite over 30 invitations being sent). Presentations for the day are found in Appendix 13. Presentation to the July 2018 stakeholder .

The purpose was to introduce SMIS to an external audience (see Appendix 14 for attendees list) and then give a demonstration of SMIS functionality, using case studies of Winter wheat; carrots and peas; and compaction and potatoes. Then the delegates were invited to test the system themselves and provide feedback of the system to the Cranfield team, using the Qualtrics survey software (See Appendix 14. Feedback from the SMIS stakeholder workshop, July 2018)

Overall, the participants were supportive and encouraging about SMIS and its potential uses to improve agricultural productivity, access big data sets and find the relationships between various variables. They liked the fact SMIS provides “everything in one place” and provides detail behind the data which can then be interrogated. There were comments that SMIS was flexible and relatively easy to use and allowed the evaluation of complex convoluted data sets quickly and easily. They liked the ability to interrogate real world data with a open query, to generate ‘what if?’ scenarios. The ability to explore correlations between yields, variable applications (agrichem & fertiliser) and other within field variations – soil texture, PCN, free living nematodes, organic matter, compaction etc. were also seen as useful.

Participants saw how SMIS could help them understand their yields better and help decision making as to future plans. They would use the outputs of SMIS to start discussions with other growers growing the same crop to find out if they have same issues as identified by SMIS. Other uses identified included: providing advice to growers; anticipating performance; checking the literature; avoiding common mistakes and pitfalls; comparing variety performance on a given type of soils; best fertilizing practices; and selecting crop varieties. One delegate likened SMIS outputs as a ‘first stage triage’, giving insight for challenges and validation of decisions.

The participants provided helpful feedback on the displays and actions when using SMIS, such as putting drop down menus into alphabetical order, improving the colour scheme and the need to ‘uncheck’ filters when running new queries. Where possible, improvements have been made to SMIS in response to these suggestions.

Other suggestions have not been addressed, often because they are beyond the scope of the present project. This included adding a ‘share button where users can offer data into the platform’; listing the insecticides used by growers (partially dictated by the level of detail

captured in Gatekeeper); incorporate satellite biomass data (but this is beyond the parameters included in the Gatekeeper database and therefore not included presently in the SMIS database). Other missing metrics that could be added at a future date include crop quality and uniformity (currently only crop quantity i.e. yield is included).

One participant wanted to compare soil types on one graph. This has not been actioned, as it was felt it was unlikely that the intended users of SMIS will be interested in different soil types: they will be growers / farmers who are based on one particular type of soil. Another participant wanted to look up the previous cropping for more than one year ago (SMIS currently can look up the 'previous [years] crop'). This would be a useful feature but the current version is unable to do this. At the moment, past crops can be queried using the 'Previous crop' filter, which applies to previous year. The user would need to repeat the query to go back further.

Participants were cautious about the uncertainty of the relationships shown in SMIS. They realised that high ranking relationships that only have a few data points need to be highlighted so conclusions are drawn with caution. This was acknowledged: The number of data items relevant to any particular query will be listed in the data spreadsheet when the options have been selected. For the Grower data the strength of the relationship is shown by the number (0-100) on the edges (lines) within the Rule Bases visualisation. The thickness of the lines for the literature data in the Rules Bases reflects the number of items and thus confidence in the results. The project team reassured the participants that if the relationship is not statistically significant, SMIS will display 'insufficient data to display' (or similar words).

One participant asked if SMIS can work on a tablet (or even a mobile phone) via a web browser. This is possible, however, because SMIS is intended for desktop web-browsers (and this is what we will perform testing and optimisation for), there are formatting issues (e.g. scaling of logos; sizing of graphs; layouts). Addressing these would require more time and effort than currently available on the project.

Looking ahead, the participants saw how SMIS can be used to set future R&D priorities. The future challenge is going to be to continue to add data to the system and to improve the efficiency of collating data. Also, in hindsight, it may have been better to start with arable data, a sector with big data sets that could have been used to build SMIS and shown the value of 'Big Data' analysis, before focusing in on the highly specialised horticultural industry, which by its nature has smaller data sets. Indeed, the Cereals and Oilseeds panel are aware of the work on SMIS (and there is a lot of cereal data in SMIS already, because of the cross rotational datasets included). There is no reason why the structure and functionality of SMIS couldn't be applied to other AHDB sectors (subject to data availability).

From the responses after the workshop (captured in Qualtrics on line), the participants' experience and satisfaction with SMIS' user interface was rated 'Good'. The speed at which the application loads was rated 'Very Good' and the output answered the questions asked well. The participants felt confident in using SMIS outputs, following the workshop. "The 'look and feel' and functionality all looked good and at an appropriate level for this sort of tool". A number of uses of SMIS were identified, including: analysing the impact of different surface tillage options; analysing the benefits of subsoiling; analysing options of increasing soil organic matter; analysing factors affecting yield; analysing options for reducing the risk of soil compaction and soil erosion.

Many of the issues raised have been incorporated within the SMIS End User Manual (Appendix 1).

### **4.3. Attendance at Grower Association and AHDB Panel meetings**

#### **4.3.1. Year 3 (2018)**

We attended the CHAPs 4-Centre Event at Harper Adams University (21st March 2018), holding discussions with Fera, John Chinn (Chair of the CHAP Board) and Frontier Agriculture regarding SMIS. On 9th April 2018, members of the project team discussed sugar beet data sets within SMIS with British Sugar, and how these may be utilised by BBRO and British Sugar. Later that month, members of the team visited Vitacress (18th of April 2018) to discuss soil management and agronomy in general, as well as SMIS and farm data management systems in particular. Rob Simmons presented SMIS to the AHDB Cereal and Oilseeds Board on 27/6/18. SMIS was discussed with several delegates visiting the Cranfield University Stand at CropTec (September 2018). Dr Simmons also presented SMIS to the Legumes Panel (R&D) of the British Pea and Bean Growers Association on 10/10/18. PGRO's invaluable assistance and inputs to the project were discussed at a meeting on 19/11/18 (with Becky Ward of PGRO and Richard Fitzpatrick of HMC peas).

#### **4.3.2. SMIS project lifetime**

Over the current life of the project, we have presented the SMIS project at over 25 horticultural meetings and have also hosted 4 stakeholder meetings at Cranfield University (June 2017; and March, June and July 2018). We have written 2 articles for AHDB levy payers' magazines.

#### **4.4. Compilation of KE events where SMIS was presented**

Please see Appendix 15. Events at which the SMIS project has been promoted

#### **4.5. Articles for AHDB distribution to levy payers**

An update on SMIS was included in AHDB's Field Vegetable Review 2018/19.

See Appendix 10. Article for the Field Vegetable Review, July 2018

## **5. Discussion**

### **5.1. SMIS as an unprecedented source of data, information and knowledge on soil management in horticulture**

The SMIS project has demonstrated that the principles of ‘big data’ can be applied to the diverse and dispersed sources of soil management data, knowledge and information in the UK horticultural sector. SMIS contains an unprecedented repository of horticultural grower data, based on Gatekeeper records (it is estimated over 40% of growers use Gatekeeper software). The database covers over 80 crop types, spanning several years (showing the cross rotational context), from a range of geographical locations across the UK. By including whole farm rotations, crops from other AHDB sectors are adventitiously included too, such as cereals, oilseeds and potatoes. Currently there are 328,890 grower data records that can be interrogated within SMIS. When accessing grower records, due diligence was paid to data protection and later, GDPR compliance.

SMIS has added significant value to this grower data by building functionality to link this to other sources of knowledge, including experimental results and written material (literature). SMIS could improve the evidence base by integrating previously diverse and dispersed sources of information and knowledge. SMIS is able to incorporate structured (e.g. Grower data), semi-structured (e.g. list of literature items, sorted by soil management issue) and unstructured (e.g. pdf documents within the literature database) data sources. SMIS provides a new and unique library of articles, papers and other literature pertinent to horticultural soil management challenges and solutions. For the first time, end users can search for information by crop, soil management issue, soil management practice and/or country, all in one place. In this way, SMIS provides a ‘one stop shop’ to advisors, growers and researchers for accessing information. The reliability of the material is inferred by whether it is quantitative, qualitative or anecdotal in origin. The integration of wide-ranging data formats and sources makes SMIS unique: such a holistic approach has never been attempted before.

### **5.2. SMIS and the application of novel agri-informatics techniques**

Novel agri-informatics approaches have been used to create, develop, operate and interrogate SMIS. Such ‘big data’ approaches are used when data sources are “too large, messy, rapid, and diverse to handle with traditional relational database management systems and statistical software programs” (McAbee et al., 2017). The variety of data and the complexities inherent in combining and analysing the data differentiates big data analytics from traditional data



analytics (Davenport, 2014). The SMIS project shows that 'big data' provides end users with much richer and more abundant information than previously available (McAbee et al., 2017).

Data visualisation techniques (e.g. SMIS 'Rule Bases') have been developed in the project to aid data representation, exploration, analysis and interpretation. Graphical representations can communicate patterns of information in succinct, yet effective ways to facilitate data exploration and inferences (McAbee et al., 2017).

### **5.3. Interrogating SMIS to better understand soil management issues**

SMIS is a powerful research tool as it addresses some of the limitations of the experimental empirical base, where conventional research projects including field trials, are subject to location and time specific variability, leading to considerable 'noise' in the outcomes. By pooling multiple, large datasets, some of this variability is inevitably 'smoothed'. Similarly, where patterns in individual grower data are often obscured by site and time specificity, the pooled dataset can unearth relationships which were previously hidden. Patterns invisible in an individual data set are more likely to be revealed and can be used as the basis for best practice in soil management. These potential benefits can't be realised from data from one business on its own or even a few businesses working together.

The extensive, integrated, unique database can be interrogated by the end user to reveal causal relationships between soil management practices (e.g. operating outside the 'workability days' window) and outcomes (e.g. yields; compaction risk). The variety of the data included in SMIS means the number of queries and combinations of factors (crop/previous crop/soil/year) that can be run is almost limitless. SMIS outputs are created by either browsing the grower database, experimental data and/or literature held within SMIS, or by running an 'established query' on the data, such as factors affecting yield, foot rot or PCN levels. This operations are described in detail in the End User Manual (Appendix 1; separate document). SMIS outputs can then be analysed to identify the drivers behind soil management problems and their solutions. The project has thus demonstrated that SMIS output can be used to inform on-farm decisions on horticultural soil management.

In terms of meaningful output from SMIS, strong, expected relationships between causes of soil management issues (e.g. use of fungicides in previous crop) and effects (e.g. foot rot index) are reassuringly found in SMIS. However, the 'big data' approach allows the end user "not to have any expectations, theories or hypotheses about the underlying relations, but rather use the observed patterns in the data to guide future decisions" (McAbee et al., 2017). Indeed, some of the observed patterns shown by querying SMIS are challenging to explain. As a specific example, one query shows that compaction risk in carrots is increased by the previous crop, notably leeks, winter wheat, savoy/green cabbage, sugar beet or spring barley.

On the other hand, when white cabbage, potatoes and mustard seed precede carrots, compaction risk was reduced. This is a good example of where SMIS raises more questions (i.e. why would growing savoy and green cabbage increase compaction risk in the following season, but white cabbage reduces it?).

This is the inevitable with the 'big data', inductive approach, where drawing inferences about underlying patterns from observations is a defining characteristic. These inferred relationships then lead to the question: "why?", prompting more investigations into the relationships shown. "Inductive methods contribute to theory development that is accurate, interesting, and testable (Eisenhardt & Graebner, 2007) as a result of its foundation in data (McAbee et al., 2017). Thus ideas of what is best soil management practice built through induction can later be tested deductively. This allows us to "hypothesize after results are known" (Kerr, 1998). In this way, SMIS can be used to set future R&D priorities for AHDB.

#### **5.4. The need for a multidisciplinary approach**

SMIS has demonstrated 'proof of concept' in using big data to manage and interpret large and varied agricultural datasets. To do this, the project has brought together field and soil scientists, soil managers, informaticists, knowledge exchange fellows, and statisticians. The success of the approach suggests new multidisciplinary skills sets are needed to develop the system architecture and functionalities (both back - and front ends) and statistical techniques further, whilst being able to interpret and translate the results into practical, field based solutions.

#### **5.5. Engagement with industry**

Engagement with industry during a series of 'hands on', interactive Workshops gave SMIS positive reviews. The participants saw the value of SMIS in accessing large, complex and convoluted datasets to reveal causal relationships between various variables, including yield, so contributing to improved agricultural productivity. They valued having all the information in one place and the ability to see 'headlines' as well as drilling down to get to the finer detail. Using SMIS was relatively easy and intuitive, and produced outputs quickly and easily. Many of the suggestions on layout and operations have been incorporated into the current version of SMIS. The stakeholders particularly liked the ability to run different 'what if scenarios' as a basis to start discussions amongst growers. A number of applications of SMIS outputs were identified, including: analysing the impact of different surface tillage options; analysing the benefits of subsoiling; analysing options of increasing soil organic matter; analysing factors affecting yield; analysing options for reducing the risk of soil compaction and soil erosion. SMIS could also provide insight for challenges and validation of decision making – referred to

as a “first stage triage” for growers in their decision making process. Looking ahead, the participants also commented that SMIS could be used to set future R&D priorities.

## **5.6. Limitations of SMIS**

Accessing grower data was a challenge. This is because of: heavy workloads throughout the year; limited time ‘in the office’ to access the data; concerns about the anonymity of the data (a robust anonymisation of the data was carried out to avoid any identification of fields, farms or individuals); 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. Some growers were reluctant to share their data (all be it anonymised) outside of their group or cooperative. They were more likely to agree to contribute data to a limited group (e.g. within one cooperative), but ironically this undermines the principles of ‘big data’ which were fundamental to the project. Although the protocol for uploading data is self-explanatory (Appendix 2), 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).

The volume of experimental data is also limited at present. To link the experimental data with grower data directly (i.e. so the same site conditions applied to be able to compare ‘like with like’) was a challenge: few entries qualified. Data from CP107B could not be related to the Grower data because of the different sites and metrics used: only 2 fields were common to both datasets and there were no common (shared) outputs (e.g. yield) or variables to predict / explain the outputs (e.g. bulk density, organic matter). The Experimental dataset was also too small to develop meaningful relationships between cause and effect (i.e. this is not ‘big data’).

Thus caution is needed regarding the current size of the database. Just because big data-sets are magnitudes larger and more complex than traditional datasets, the statistical relationships still have to be significant before valid conclusions (and decisions) can be drawn, especially when these outcomes are related to farmer practices and the consequence of using them. McAbee et al. (2017) warn “drawing strong conclusions from relatively few observations is probably not the best application of inductive methods”. As an example, the performance of individual crop varieties identified by SMIS should not be taken as any kind of official endorsement or embargo. 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.

It is recognised that currently, the grower database originates primarily from Gatekeeper records. Automatic uploading of other datasets (e.g. Muddy Boots) could not be developed, as access to the Muddy Boots API was not granted. Also, current data exchange files are

likely to change format (requiring new programming in SMIS for the future version). Finally, grower records were in pdf format, requiring time consuming, manual uploading of data into SMIS.

The certainty and confidence in the rules and relationships developed within SMIS (and any decisions based on them) depend on the extent and quality of the database. Sourcing data from Gatekeeper records assumes an acceptable quality of input data. However, it should be remembered that the data was not collected as part of a research project with a robust experimental design. Nor was it collected in order to develop a soil management database, as is clear from the data 'headings' which can be vague (e.g. 'applications'), rather than describe a specific soil management practice. Another implication of this is that 'double counting' of operations may take place when growers enter data (e.g. 'number of field operations' and 'applications' could be the same activity).

Given the age of some records, it was not possible to clarify what these generic operations actually were. This point also explains why some queries could not be run, because of 'missing metrics'. For example, there are no metrics of crop quality (only quantity, i.e. yield). In some cases, proxy metrics can be used (but with caution). For example, incidence of soil compaction is not recorded (nor even indicators of soil compaction, such as bulk density). Therefore 'proxy' indicators of compaction (e.g. where subsoiling operations had been undertaken) had to be used instead. There are limitations with this: the records do not reveal whether in reality the subsoiling was needed to remediate existing compaction or was to used as a preventative measure.

SMIS can work on a tablet (or even a mobile phone) via a web browser. However, because SMIS is intended for desktop web-browsers (and this is what we will perform testing and optimisation for), there are formatting issues (e.g. scaling of logos; sizing of graphs; layouts). Addressing these would require more time and effort than currently available on the project.

## **5.7. Final outputs from the project**

The final outputs of CP107D are

- Annual Reports (2016, 2017 and 2018)
- Documentation on SMIS back end and front end (See Appendices)
- SMIS End User Manual (Appendix 1)
- SMIS system itself (First Release Candidate [RC] sent to AHDB September 18; final version October 2018). This includes the reference / HTML documentation included with the source code and installation script on the AHDB server.

## 6. Conclusions

The aim of the SMIS project is to provide comprehensive and coherent information to the horticultural sector in support of decisions on sustainable soil management. The intended outcome of this data-driven decision-making is to improve crop productivity, whilst reducing the causes and symptoms of environmental damage such as soil compaction.

- The need for integration of data, information and knowledge on soil management in horticulture was identified in a grower survey back in 2013.
- SMIS has added significant value to grower data by building a functionality that links this to other sources of knowledge, including experimental results and written material. By integrating previously diverse and dispersed sources of information and knowledge, SMIS has improved the evidence base. The integration of wide-ranging data formats and sources has never been attempted before.
- The SMIS project shows that ‘big data’ approaches can be used to interrogate the comprehensive and coherent database to provide the horticultural sector with much richer and more abundant information to support decisions on sustainable soil management than previously available.
- For the first time, end users can search for information by crop, soil management issue, soil management practice and/or country, all in one place. In this way, SMIS provides a ‘one stop shop’ to advisors, growers and researchers for accessing information. The number of queries that can be run in SMIS is almost infinite, given the number of ‘soil/crop/previous crop/soil type/year’ combinations in the dataset, but could include questions such as: What factors affect crop yield? What factors lead to soil compaction? What are the impacts of carrying out operations outside of ‘workability’ windows (i.e. when the soil is too wet)? How effective are fungicides on PCN levels?
- New and emerging methods of data management and processing (‘agri-informatics’) allow meaningful interpretation of large datasets to unearth patterns undetected before. Novel informatics techniques are used to create and then extract patterns of ‘cause and effect’ regarding soil management practices (and their outcomes) in different scenarios (e.g. soil type, crop, previous crop, year, etc.).
- However, accessing grower data was a challenge. Caution is needed regarding the current size of the database. The statistical relationships still have to be significant before valid conclusions (and decisions) can be drawn. 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.
- The outputs of SMIS can be used to a) reaffirm current understanding of the effects of soil management practices and b) unearth new insights of possible causes of soil management

issues and effects of soil management practices (requiring new research to validate those outputs). In this way, SMIS can be used to set future R&D priorities for AHDB.

- Industry representatives see the value of accessing large, complex and convoluted datasets to reveal factors affecting agricultural productivity. They value having all the information in one place and the ability to see 'headlines' as well as drilling down to get to the finer detail. Stakeholders find using SMIS is relatively easy and intuitive, and produces outputs quickly and easily. They particularly like the ability to run different 'what if scenarios' as a basis to start discussions amongst growers. SMIS provides a "first stage triage" for growers in their decision making process.
- The SMIS interactive platform provides AHDB-Horticulture, and its growers, agronomists and land managers insights of contextual, effective soil management practices that can inform development of advice and guidance. Ultimately, the beneficiaries of SMIS will be farmers and growers.

## 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.

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.
ES2016	ECMAScript 2016
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
HTTP(S)	Hypertext Transfer Protocol (Secure)
JSON	JavaScript Object Notation
LandIS	<p>Land Information System (LandIS; <a href="http://www.landis.org.uk/">http://www.landis.org.uk/</a>) is a soils-focussed information system for England and Wales. <a href="http://www.cranfield.ac.uk/About/Cranfield/Themes/AgriFood">http://www.cranfield.ac.uk/About/Cranfield/Themes/AgriFood</a></p> <p>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.</p>

	The Cranfield <a href="#">Soil and AgriFood Institute (CSAI)</a> , 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
NoSQL	Non-SQL / Non-Relational
ODM	Object Data Modelling
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 ( <a href="https://www.w3.org/RDF/">https://www.w3.org/RDF/</a> and <a href="http://www.w3.org/TR/rdf11-concepts/">http://www.w3.org/TR/rdf11-concepts/</a> ), 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
SCSS	Sassy Cascading Style Sheet
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.
SMIS AT	SMIS Analytics Toolkit
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
URI	Uniform Resource Identifier



VESS	Visual Evaluation of Soil Structure
VM	Virtual Machine
VSA	Visual Soil Assessment
W3C	The World Wide Web Consortium ( <a href="https://www.w3.org">https://www.w3.org</a> ), owners of the RDF schema.
XML	Extensible Markup Language

## 8. References

Davenport T.H. (2014). *Big data at work: Dispelling the myths, uncovering the opportunities*. Harvard Business School, USA (2014)

Eisenhardt K. and Graebner M. (2007). Theory building from cases: Opportunities and challenges. *Academy of Management Journal*, 50 (2007), pp. 25-32

Kerr N. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, 2 (1998), pp. 196-217

McAbee, S.T., Landis, R.S. & Burke, M.I., (2017). Inductive reasoning: The promise of big data. *Human Resource Management Review*, 27(2), pp.277–290. Available at: <http://www.sciencedirect.com/science/article/pii/S1053482216300377>.

Rickson, R.J. and Deeks, L.K. (2013). A gap analysis of soil management research and knowledge transfer in horticulture to inform future research programmes. Final report to the Horticultural Development Company. 64pp. NSRI, Cranfield University, UK.

## **9. List of appendices**

Appendix 1. SMIS end user manual

Appendix 2. Protocol for extracting data from GateKeeper

Appendix 3. SMIS project Gantt chart summarising milestones and deliverables

Appendix 4. SMIS database architecture technical documentation

Appendix 5. SMIS system architecture

Appendix 6. SMIS use case documentation

Appendix 7. SMIS web interface report

Appendix 8. SMIS technical documentation and user manual report

Appendix 9. SMIS Steering Group

Appendix 10. Article for the Field Vegetable Review, July 2018

Appendix 11. Presentation to the June 2018 stakeholder workshop

Appendix 12. Feedback from the SMIS workshop, June 2018

Appendix 13. Presentation to the July 2018 stakeholder workshop

Appendix 14. Feedback from the SMIS stakeholder workshop, July 2018

Appendix 15. Events at which the SMIS project has been promoted