Towards data democracy in digital agriculture

Peter Dahlhaus[1,2], Helen Thompson[1] and Andrew MacLeod[1]

[1] Centre for eResearch and Digital Innovation, Federation University Australia, Ballarat, VIC 3350, www.cerdi.edu.au
[2] Corresponding author: p.dahlhaus@federation.edu.au

Abstract

The present era provides unprecedented volumes of data that could be available to decision models and tools that can be used to grow the production and profitability of Australian agriculture. The challenge is in providing equitable access to all the available data – public and private – in a seamless manner, regardless of its disparate custodianship and collective heterogeneity. Implementing data democracy requires both the technology to interoperably federate data as well as the motivations for custodians to provide their data for all to benefit. A research project being undertaken in the Corangamite region in Victoria is successfully implementing a model in which numerous soil health data sets owned and managed by a variety of custodians are brought together in a single web-portal. A number of use-cases were developed with end-users to identify the required web-portal functions. Data contributors are rewarded with a greater understanding of their soils both historical and present status, and the wider community, which includes land managers and other stakeholders, have access to increased soil health information that enables more informed decision-making in catchment management activities across the region.

Keywords

Soil health, data interoperability, community contributed data, soil data, crowdsourcing, eResearch.

Introduction

Our present time period, which is sometimes referred to as the New Digital Age or Era of Big Data, provides unprecedented opportunities for a deeper understanding and appreciation of global environments, including agricultural and agribusiness environments. In the last decade the volume of digital data relating to agricultural landscapes has grown exponentially, much of it having been collected by sensors (Keogh and Henry 2016; Stubbs 2016). In addition, data availability has vastly improved as governments in many countries – including Australia – adopt open data policies (Productivity Commission 2016; Welle Donker and van Loenen 2017).

Concurrent with Big Data emergence has been a move around the world towards deceased government size and steadily adoption of economic liberalisation policies such as privatisation. As a result there has been a shift in data collection from the public sector to the private sector: more agricultural data is now being collected outside of government agencies than within them. The collection of data in the private sector comprises both on-farm and off-farm data, and includes everything across the entire value chain from soil tests to consumer purchases. The result has been an ever-growing plethora of chaotic data collections, most of which are invisible to decision makers in both the private and public spheres. Even where open data is provisioned, it is often in formats that are intelligible only to experts with the knowledge and technology to enable interpretation (i.e. data aristocracy). In some cases, custodians control or restrict the use of data (i.e. data dictatorship) or selectively distribute data based on the perceived need of access (i.e. data oligarchy). In other cases, data is freely distributed or shared in an *ad-hoc* fashion, with users creating their own data sets by combining whatever data they can access (i.e. data anarchy). Australia's agricultural production and profitability would vastly improve if everybody has equitable, timely access to all the data required to answer the frequently asked use-case questions in both the private and public sectors (i.e. data democracy).

However, simply collecting more data and increasing its availability does not necessarily answer the end-users' questions. Indeed, although technology and policy have enabled some ready access to more data, there remain limitations in how to transform this data into improved decisions to aid agricultural production and profitability. Besides the limitations experienced by the user in discovering, accessing, comprehending and harmonising the data, few users have the capacity or desire to undertake the task. Supplying the metadata that informs the user of the data collection methods, veracity and resolution are of critical importance. Recent assessments of agricultural decision support tools found that common limitations in system models for decision support are: 1) data scarcity (quantity, resolution and quality), and 2) inadequate knowledge systems to effectively communicate the results to the end-user. These limitations are greater obstacles to use of the tools than gaps in theory or technology (Capalbo et al. 2017).

Data interoperability

Current technology can assist by making disparate data interoperable – that is – making the data usable in a seamless manner regardless of its original collective heterogeneity. Data interoperability is accomplished by the transformation of data, most commonly by using a standard, or a representation readily understood and utilised by both the data supplier and consumer. International interoperability standards are often derived through global collaborations of experts who co-develop and maintain them (e.g. OGC 2017).

A data network consists of data sources that are managed autonomously and that typically have heterogeneous structure and content. Data networks utilise spatial data infrastructure to federate the data provisioned by the disparate custodians. An example is provided by that used at the Centre for eResearch and Digital Innovation (CeRDI) at Federation University Australia (FedUni), as illustrated below (Figure 1).

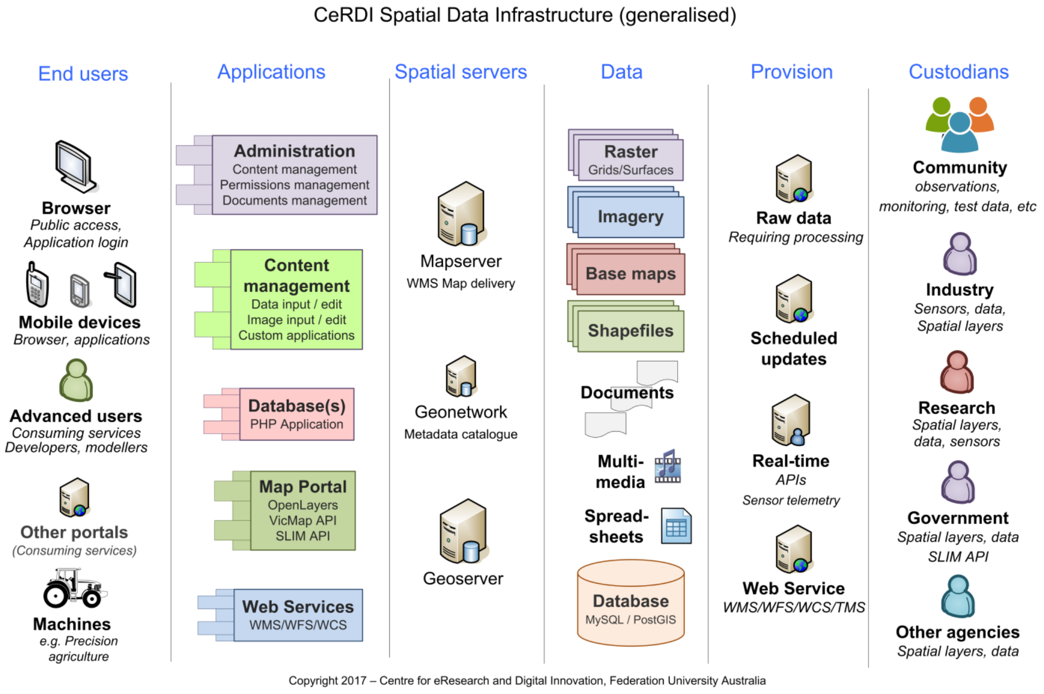

Figure 1. An example of the spatial data architecture used to interoperably federate data.

Data is provisioned preferably via real-time access to the remote databases by integrating the interoperable web services they each provide. Since not all custodians are capable of serving data in real time, scheduled updates are sometimes used where data is provisioned via a public access portal (e.g. data.gov.au) or simply via a cloud-based service on a regular basis. Where data is provisioned by community groups and individuals, it may be raw data that is digitised and structured into a web-based database to which the custodians are given access (via a login) for on-going data management and curation.

Provisioned data may take any form, and is federated via a suite of open-source software tools generally referred to as a spatial information services stack (Golodoniuc et al. 2011). The tools are built upon software projects fostered and supported by the Open Source Geospatial Foundation (www.osgeo.org). The federated data can then be viewed via a web-browser, with the portal interface built to bespoke requirements upon the foundations of the OpenLayers (openlayers.org) javascript library. Other javascript libraries are leveraged to provide user-interface components and functionality.

Data democracy using soil data

The Corangamite Soil Health Knowledge Base is a novel collaborative research project between the Corangamite Catchment Management Authority (CMA) and CeRDI. The aim of the research is to develop a comprehensive, informative, intuitive-to-use knowledge base of soil health information that will assist the broader community to respect the values of the soils of the Corangamite region of south west Victoria. The

Soil Health Knowledge Base (www.ccmaknowledgebase.vic.gov.au/soilhealth/) has two main components: (i) a searchable eLibrary of digital documents, webpages, images and multimedia; and (ii) an interactive map portal to discover spatial soil data. The data and information portal is based on the spatial data infrastructure shown in Figure 1 and won the 2015 Victorian Spatial Excellence Award and the 2015 Asia Pacific Spatial Excellence Award in the Environment and Sustainability category.

The Soil Health Knowledge Base research and development project has adopted a participatory approach focussed on answering the community's frequently asked questions on soil health. The project is built around use-cases such as: "a soil scientist using a mobile tablet in the field can see a summary of the soil characteristics and soil properties at their location derived from the private and public data available within a set radius within a soil series unit." and "a farmer can spatially see their farm on a map portal and investigate the changes in soil nutrition values in a selected paddock over time." The project thus aims to pilot the use of both public open data and private data contributed by the community. The intention is to test the use-cases with equitable access to all the available data, i.e. to establish a data democracy. However, private soil test data can be sensitive, and may affect factors such as property real estate values. Therefore, farmers are provided with a login to allow private access to applications and functions that are unavailable in the public view. As an incentive to contribute data, farmer logins also enable them to view time series graphs, animations and other visualisation tools that show trends in soil nutrition parameters over time (an example is shown in Figure 2), thus improving their knowledge and understanding of soil health and limitations in their paddocks.

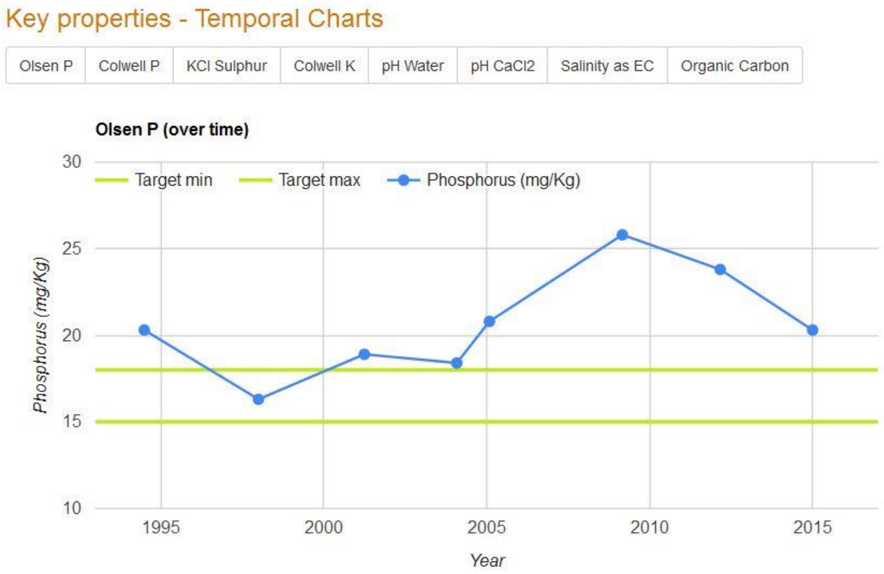

Figure 2. An example function that tracks soil nutrient data over time for an individual paddock: the tabs at the top display equivalent data for other nutrients.

To date this has encouraged around 95 farmers to contribute soil tests taken over a period of up to 25 years. The soil test data are used to aggregate the values for more stable soil parameters for the public view without disclosing the actual locations. For example, a report for any location may be "there are 10 soil tests within this unit within a 1 km radius, with the following average values ... Bulk Density, Organic Carbon, Phosphorus Buffering Index, Sum of Cations, Soil Texture, etc.". In this way private data can contribute to the overall knowledge of soil health for all to access within the catchment, without disclosing test sites. Where no private soil data has been contributed, the reporting tool defaults to the next best data source, such as State Government digital soil maps or the Soil and Landscape Grid of Australia, both of which have much greater error around the modelled data.

Conclusions

In agriculture, eResearch is revolutionising many traditional areas of research, such as soil mapping (e.g. Sanchez et al. 2009), agricultural production (e.g. Whelan and Taylor 2013), food security (e.g. Gebbers and Adamchuk 2010) and biotechnology (e.g. Sansone et al. 2012) to name a few. A deeper appreciation and understanding comes from the sheer volume of information that can be amassed about any particular topic or place in the landscape, and how that information can be dynamically synthesised to provide the most definitive answer to the user's question.

Adoption of precision agriculture (e.g. yield maps), sensors (e.g. soil moisture and temperature) and remotely sensed data (satellite, aerial, ground) in digital agriculture is growing rapidly. Future eResearch and developments for the Soil Health Knowledge Base will focus on how to interoperably federate the new data, harmonise it and feed it into dynamic models that will provide timely decision support for agriculturists and catchment managers to protect, enhance and restore soil health.

References

Capalbo SM, Antle JM and Seavert C (2017). Next generation data systems and knowledge products to support agricultural producers and science-based policy decision making. Agricultural Systems (in press). (doi.org/10.1016/j.agsy.2016.10.009).

Gebbers R and Adamchuk VI (2010). Precision agriculture and food security. Science 327 (5967), 828–831. (doi:10.1126/science.1183899).

Golodoniuc P, Fraser R, Woodcock R and Wyborn L (2011). The Spatial Information Services Stack (SISS). In: 5th eResearch Australasia Conference, Melbourne, Australia, 6–10 November, 2011. (https://publications.csiro.au/rpr/download?pid=csiro:EP125738&dsid=DS3).

Keogh M and Henry M (2016). The implications of digital agriculture and big data for Australian agriculture. Australian Farm Institute. 84p. (http://www.farminstitute.org.au/publications-1/research-reports/the-implications-of-digital-agriculture-and-big-data-for-australian-agriculture).

OGC (2017). Agriculture DWG. Open Geospatial Consortium, Agriculture Domain Working Group. (http://www.opengeospatial.org/projects/groups/agriculturedwg).

Productivity Commission (2016). Data availability and use. Productivity Commission Draft Report. Australian Government, Canberra. 652p. (http://www.pc.gov.au/inquiries/completed/data-access/thedraft).

Sanchez PA et al. (2009). Digital soil map of the World. Science 325 (5941), 680–681. (doi:10.1126/science.1175084).

Sansone S-A et al. (2012). Toward interoperable bioscience data. Nature Genetics 44 (2), 121–126.

Stubbs M (2016). Big data in US agriculture. Report prepared for Members and Committees of Congress. Congressional Research Service. 17p. (https://fas.org/sgp/crs/misc/R44331.pdf).

Welle Donker F and van Loenen B (2017). How to assess the success of the open data ecosystem? International Journal of Digital Earth 10 (3), 284–306. (doi:10.1080/17538947.2016.1224938).

Whelan B and Taylor J (2013). Precision Agriculture for grain and production systems. CSIRO Publishing, Melbourne, VIC. 200p.