

The rise of information science: a changing landscape for soil science

Pierre Roudier¹, Alistair Ritchie², Carolyn Hedley¹, David Medyckyj-Scott¹

¹ Landcare Research, Manaaki Whenua, Private Bag 11052, Manawatu Mail Centre, Palmerston North 4442, New Zealand

² Landcare Research, Manaaki Whenua, PO Box 69040, Lincoln 7640, New Zealand

E-mail: roudierp@landcareresearch.co.nz

Abstract. The last 15 years have seen the rapid development of a wide range of information technologies. Those developments have been impacting all fields of science, at every step of the scientific method: data collection, data analysis, inference, science communication and outreach. The rate at which data is being generated is increasing exponentially, giving opportunities to improve our understanding of soils. Parallel developments in computing hardware and methods, such as machine learning, open ways to not only harness the “data deluge”, but also offer a new way to generate knowledge. Finally, emerging data and information delivery protocols are leveraging the outreach power of the World Wide Web to disseminate scientific data and information, and increase their use and understanding outside the boundaries of a given scientific field. However, the nature of this data is mostly new to soil science, and requires adaptation to its diversity and volume. In particular, the integration of the significant amount of legacy soil data collected throughout decades of soil science can be problematic when all necessary metadata is not available. Likewise, knowledge accumulated by our scientific field needs to be acknowledged by - rather than opposed to - numerical methods. While the introduction of this set of emerging technologies is enabling soil science from different points of view, its successful implementation depends on the ability of soil scientists to act as knowledge brokers and support numerical methods.

1. Introduction

The explosion of information technologies has been the most important development in science over the last 15 years. In particular, it has permitted data to be generated at higher rates than ever before and keeps on increasing exponentially. This phenomenon has been described as the “data deluge”. It has had a considerable impact on science in general: whole disciplines have moved from data-poor to data-rich in a significant shift from a situation where the collection of one data unit is more costly than its analysis, to the opposite situation. Soil science is no exception. Further, as technologies underlying the data deluge develop and mature, they will continue to change the way soils are studied.

Simultaneously, analysis capabilities have been growing to support the exponential increase of data being created [1]. Computing hardware underwent drastic developments in the last decade enabling the analysis of the increased flow of data. Key technology advances have dramatically lowered the cost of storage, amplified by the use of commodity hardware, enabling modern, workload-driven storage of data that is flexible and easy to scale. The invention of high-performance computing (HPC), servers linked up to behave as one, and parallel programming facilitates the use of high numbers of computing cores to solve complex problems [2]. Finally, overarching the increased amount of available data and the development of high-performance computing hardware, the research field of



artificial intelligence has produced different numerical methods that are able to handle the large and varied datasets becoming available. Amongst these, machine learning are prediction methods that induce new properties from existing data. Related to machine learning methods, data mining are methods to discover new knowledge from exploring relations in the existing data.

The implications of these technological developments for science are truly disruptive. Pioneering computer scientist Jim Gray has been formulating this as a new paradigm for science exploration and discovery: the fourth paradigm. According to this new scientific paradigm, technologies would lead to the collection, analysis and visualisation of increasingly large amounts of data that would change the very nature of the scientific discovery process [3]:

The world of science has changed, and there is no question about this. The new model is for the data to be captured by instruments or generated by simulations before being processed by software and for the resulting information or knowledge to be stored in computers. Scientists only get to look at their data fairly late in this pipeline. The techniques and technologies for such data-intensive science are so different that it is worth distinguishing data-intensive science from computational science as a new, fourth paradigm for scientific exploration.

Physicists, astronomers, biologists, chemists, earth and social scientists are all benefiting from access to the tools and technologies that integrate “big data” technologies into standard scientific methods and processes. This paper will discuss both the opportunities and the risks that are associated with the growing influence of these technologies in soil science.

2. The impact of big data technology developments in soil science

2.1. *The data deluge in soil science*

The “data deluge” is a term coined to describe the acceleration of the production of data over the last decade. The structural changes brought about by the data deluge have originally been described using the three “Vs” [4]:

Volume: the amount of data generated is very important, and challenges the way data has been processed using traditional approaches. [2] reported that the world's information storage capacity grew at an annual rate of 25% between 1986 and 2007 to support the acceleration of data generation, but the amount of information created exceeds the world's storage capacity [5, 1],

Variety: the sources of scientific data, and the form into which it is formatted, are increasingly variable and diverse. The collected data can either be structured or unstructured. This requires tools that are able both to cope with increasingly multi-dimensional problems, and to interact with different sources of data and combine these into a useful set for further analysis,

Velocity: the speed at which data is being created is faster than ever before, due in particular to the increased availability of various sensors. The ever-increasing flow of data requires systems that can support updates on the data on which its analysis rely.

A fourth “V” has been added since this original formulation:

Veracity: refers to the problems of data quality, and to the various levels of uncertainty associated with each of the datasets used to support a decision. Data adequacy, a concept integrating the uncertainty of the dataset and its relative relevance to answer a given question, needs to be assessed on a case-by-case basis.

The terms most widely used to describe the evolution of data (“big data”, “data deluge”) are unfortunate: what characterises the emerging kinds of data is more than just its volume, the various forms it takes or the rate at which this data is streaming. What makes it really new is its connectedness, and the fact that the value associated with these datasets is substantially increased when combined [6]. The four “Vs” are an opportunity for new insights into soil science: fostered collaborations, scaling-up of scientific concepts by patching together dispersed datasets, meta-analysis of previously published works, discovery and modelling causality in an ensemble of complex datasets, real time analytics and processing.

All these aspects represent a significant challenge to soil science. Traditionally, soil data, because it is generated from field observations and further lab analysis, is scarce and costly [7]. These measurements are very demanding, both in time and associated costs, which is why the number of locations sampled is usually low. But the development of sensing methods proposes an alternative type of soil data: soil properties can be observed indirectly, but at a much higher rate, both in time and space, and for a fraction of the price per sample.

An important part of the data deluge is generated by sensing environmental conditions. An example of such developments is proximal soil sensing (PSS). Proximal soil sensors enable field analysis of soil attributes, such as soil carbon and moisture, and on-the-go sensor surveys can define the spatial variability of these soil attributes. Further, wireless sensor networks can be deployed to monitor the temporal change of the spatially defined soil attributes. These tools make it possible to map and monitor changes of soil attributes in the soil profile at a resolution, both in space and time, that was previously impractical. They also enable a scientist to measure those soil properties in situ, with limited disruption of the soil. A review of the ever-growing array of sensors available for soil applications is provided by [8]. Another example of developing data available for soil scientists to use is remotely sensed environmental data such as LiDAR for high-resolution elevation models, or satellite missions such as Landsat or ASTER [9].

The form that the flow of incoming data takes is very different from the traditional soil datasets, and requires the scientist to adapt and make more robust the traditional data management methods. The use of emerging data in soil science comes with a trade-off: the on-the-spot accuracy of models using these data sources is generally lower than the traditional lab measurements, but is able to capture a greater part of the important spatial variability observed in soils because more locations can be sampled.

2.2. Development of analysis capabilities to address data growth

The technological ability to process data has been growing in parallel with data volume and velocity [2]. This rapid development has been two-fold: computing hardware on one hand with high-performance computing (HPC) platforms making use of a number of cores; and machine learning methods on the other hand making the most of the HPC capability and providing mathematical tools that can investigate complex systems using quantitative methods and extract information from all this data. This parallel development fostered a new paradigm for science: using data mining techniques, scientists can try to discover new patterns from the data itself, rather than formulating a hypothesis and testing it empirically [10]. Such approaches have been successfully applied in soil science, and pedotransfer functions (PTF) are generated by interrogating important soil profile databases [11].

Environmental data, with satellite missions such as Landsat or SRTM, are providing soil scientists with environmental data at a very good spatial and temporal resolution. As a consequence, scientists are provided with prior information on the areas they may wish to study, and such environmental covariates can be used to inform their sampling design [12]: for example, the sampling locations at which soil spectra should be taken, or at which the soil moisture sensors should be installed, could be determined by interpretation of a stack of environmental data layers.

Digital soil mapping [13] is another application of machine learning methods in soil science: soil profile data, whether generated using traditional lab methods or using sensing methods such as visible near-infrared spectroscopy, can be combined with environmental covariates to predict soil attributes over the landscape [14], or to improve the spatial resolution of existing data [15]. On top of generating new (spatial) information, data mining has also proved to be a useful tool for knowledge discovery from soil maps derived by soil surveyors [16], or from analyzing large soil databases [17]. In this instance, these methods provide tools to investigate the very complex interactions between soil properties, which are difficult to measure, and a range of widely available environmental covariates derived from proximal or remote sensing technologies.

The application of quantitative methods to soil science, either to predict new soil properties from measured ones (PTF) or to map given soil properties in space (DSM), is not new and has been

occurring in our field for the last two decades at least. However, the development of the computing power necessary to process the ever-increasing amounts of available environmental data is occurring on alternative platforms for soil scientists. As computing power is developing, the amount of data to be processed has kept up well with Moore's law. As a consequence, while most of the computing tasks have traditionally been done on single core desktop computers, computing efficiently now requires to (i) move to high-performance, multi-core cluster computers, and (ii) to do the processing task close to the data to limit the input/output overheads associated with moving big chunks of data around. To enjoy the benefits of high-performance computing, soil scientists need to adapt their analysis and tools, and change the way they currently store and distribute data.

3. Challenges and opportunities

3.1. *Keeping research reproducible*

The concept of reproducibility is at the very heart of the scientific method [18]. The ability for scientific results to be reproduced by peers is the main method to confirm the validity of those results. To achieve this, data and methods need to be accessible to peers. Moreover the scientific method is based on an incremental process: new results generally build on top of previous results, so it is important that researchers have the ability to access and use the methods developed to obtain the state-of-the-art results to generate new science. On top of guaranteeing the necessary reproducibility of the claimed results, data and model availability are working together to accelerate scientific progress and upscale [19].

Because the technical implementation of modern machine learning tools is impacting on the analysis they are supporting, their use is a challenge to research reproducibility. Beyond the parametrisation of the method, there are different ways to implement a given machine learning algorithm [20]. In the past, both data and methods have been reported in scientific papers. The rise of numerical approaches based on various and voluminous data sources and running on HPC platforms is challenging the way science reproducibility has been ensured traditionally. In its current format, the “material and methods” section of scientific papers does not allow for reproducibility, because data has grown much too large to be included, and because the implementation of modern machine learning tools is impacting on the analysis they are supporting [21]. To ensure reproducibility and allow the fostering of science by sharing not only algorithms but also their software implementation, several authors are making the case for sharing code in addition to data.

3.2. *Open soil science*

In line with these developments, empirical data suggests there is a significant trend in scientific journals to make not only data, but also code available alongside papers published in their columns [20, 22, 23]. Beyond reproducibility, making data and code available are a way to accelerate the pace of research, and improve its cost-efficiency [24]. This idea is formalised by the concept of open science, which promotes open access of science articles along with the availability of code and data [18]. Open science is a virtuous circle where available data can improve models, available models can improve data collection, and open access publication spread ideas further and faster, thus improving the returns on the science endeavour.

Solutions have been proposed. Coming from the software development community, public code repositories like Github now support this evolution giving the opportunity to associate a digital object identifier (DOI) to a code repository [25]. A DOI is a way to uniquely identify an electronic object and its metadata, allowing it to be formally cited in a journal article, accessed by other researchers and recognised as a piece of research. For data, on-line repositories such as Figshare and Zenodo also allow the publication of any research outputs, such as datasets, as a citeable object in its own right.

3.3. *Integrating past, present and future data*

By their nature, data-oriented approaches often weave various sources of data together. Data availability and structure are guiding the pace at which such science can progress, which in turn raises the value of datasets that have been collected in the past. The emerging sensing techniques, for example, are not exclusive to traditional approaches: reference data is indeed needed to calibrate most quantitative, sensor-based approaches. Legacy lab data can play this role, but its harmonisation and integration with sensing data can be problematic depending on the state of the associated metadata. Unfortunately, scientific data has traditionally been managed on a per-project basis, making data re-use a very challenging task. A consequence of this is “dark data” – data that has been collected but has become invisible to scientists because it is inaccessible [26]. Shining a light on “dark data” by making it available will ensure it can be used for further applications, improving the cost-efficiency of such datasets. It is an opportunity for soil science to give a second youth to the significant amount of legacy data that has been collected over time by soil scientists [27].

The rescue of soil science's data legacy has been one of the major drivers of the GlobalSoilMap project [28]. But while making legacy datasets available in a digital form is necessary for them to be distributed and used in coordination with other existing data, it is not enough, legacy data needs to have metadata associated to it, and its quality needs to be assessed and raised to modern standards [27]. Additionally, when collating several soil datasets, it is also necessary to make them compatible using harmonisation techniques [29, 30]. Numerical techniques are being developed to do so [31], but the use of semantic mediation techniques, such as data mapping and linking based on common information concepts and controlled content (e.g. ontologies, taxonomies, glossaries, linked data, etc.) needs to be investigated in order to facilitate data integration more systematically.

3.4. *From sensors to analysis: using interoperability specifications*

The drive towards the fourth paradigm requires researchers to make their data publicly available, but also to improve the curation of past datasets to improve their cost-efficiency and allow new results to be derived even after the project is finished. Data-rich approaches underline the need to extend the life cycle of the datasets, and to foster their re-use. While data curation is a new task to add to the soil research departments, those technologies also provide solutions to share and distribute data and methods used in science.

Making data available online is a challenging task. Data needs to be presented in a standard form, and associated with its metadata, so it can be linked with other existing datasets to be truly useful [32]. A set of tools to do so is provided by the data modeling methods, which were first introduced in the domain of software development. Data modeling methods help formalise the scientific concepts and objects that scientists manipulate, along with their relationships. These data models explicitly define the structure of data. In soil science, the ANZSoilML initiative [33] aims to model soil science data from the pedologists perspective. The information model describes and organises soil data, and also provides for the capture of necessary metadata, ensuring the data generation process is documented, and reproducible. Data models like ANZSoilML, if adopted by the community, also ensure the interoperability of the various data sources, which lowers the cost of the data collation for the scientist.

Such data models can then be implemented in an operational solution using web services. These are standardised protocols to stream data over the Internet rather than store it in downloadable files. The advantages of web services over static files are that they allow data to be more dynamic, providing users with updates as data change. Web services also allow the use of new tools, beyond the traditional spreadsheet on a desktop computer, facilitating the weaving of various data sources on various computing platforms (desktop, cloud, but also mobile). To facilitate data sharing, the use of existing standards (such as the Open Geospatial Consortium web services), when appropriate, is encouraged [34].

3.5. *Adapting the analysis workflow*

In order to keep up with the increasing amount of data generated, HPC and machine learning methods have been very active fields of research -- even going beyond expectations by proposing a fourth paradigm for scientific discovery based on data [3, 10]. The trade-off from these new tools is that researchers need to embrace specific tools to benefit from the technological breakthroughs they provide. An example of such trade-off is that HPC platforms do not generally come with a graphical user interface: the analysis is sent to the computing cluster as a script (i.e. a high-level programming language that is interpreted by another program at runtime rather than compiled by the computer's processor as other programming languages such as C or C++).

While the learning curve associated with these tools might be steep for some, the input of soil scientists into this process is crucial. First, domain knowledge is necessary to ensure sensible use of those technologies, i.e. that they not only provide results that are compatible with the existing knowledge base in soil science, but they also answer valid questions. Moreover, the end goal of science is to produce knowledge, as opposed merely to turn data into information. The role of the domain expert to this regard is two-fold. It is to aggregate the information patterns detected by machine learning methods with the existing knowledge base so they can be turned into knowledge itself and be used to build on; it is also to ensure that the staggering amount of knowledge (i.e. largely undocumented experience, insights, knowledge and skills) accumulated throughout decades of soil science in the lab and in the field is acknowledged and used. Numerical data can benefit from it and support the generation of new knowledge, but not replace the soil scientist's domain expertise. In this sense, the input of soil scientists as knowledge brokers for the successful implementation of data-oriented methods.

3.6. *Opportunities in soil science outreach*

The contribution of information technologies to the science endeavour is not limited to data generation and analysis; innovations have also taken place in how to share and present data. The emergence of the World Wide Web as the main communication platform and the development of mobile technologies are examples of this. Developments in web technologies allow advanced software applications to be used in many different computing environments as opposed to traditional desktop-side technologies (an example of this is web mapping technologies such as Google Maps, which have popularised maps as a form of information presentation, previously limited to user-unfriendly desktop GIS environments). Web data services, such as OGC web services, offer a standardised way to “stream” data over the web. Those services can connect a data provider to a range of data consumers, inside or outside the research community, but can also connect the data provider directly to processing platforms. The agility of data services technologies can improve soil data dissemination, and improve the return on investment of soil data but maximising data re-use. Moving data provision on the server side does however make it vulnerable to problems associated with cloud storage, such as unexpected server downtime. It also raises synchronisation needs for offline applications, a common case when working in the field. These technologies are changing the way science is shared across scientific communities and delivered to policy makers and how science interacts with the general public, by providing new tools and the communication channels to do so. In particular, mobile devices and data visualisation techniques can build upon those data streams, and tailor the way science can be consumed by a variety of stakeholders [35].

Information technologies are also fostering and enabling feedback loops from the stakeholders to the researchers. Analysing the logs of web sites and web services can assess whether scientists are or are not answering the questions asked by the public. Additionally, most smartphones carry a GPS chip, a colour high-resolution CCD camera and full-time internet connection, can be used as sensors [36], and even contribute to the data collection effort [37] or validation of data.

4. Conclusion

Disruptive developments in the field of information technologies impact on all aspects of science, and soil science is no exception. With the contribution of technologies such as remote sensing, proximal soil sensing and wireless sensor networks, the data used to improve our understanding of soil is becoming more intensive and more dynamic. It is also more diverse, and while the number of sources of soil data is increasing, it is important to combine these to create potentially useful soil information. The integration of this emerging stream of data with existing soil data is also important in order to leverage and acknowledge the data collected over the last decades using traditional approaches. The data management, in particular the available metadata of these legacy datasets, will have an impact on the feasibility of this integration. High-performance computing hardware, machine learning and data mining methods are supporting these important changes in data, which allow us to maximise the value of our data. The combination of data richness and HPC hardware enables numerical approaches in soil science, previously impractical, based on the recent developments in machine learning methods.

However, it is important to notice that the challenges raised by the “data deluge” are not so much the amount of data per se, but the novel analysis to produce knowledge and help insightful decisions [38, 1]. While the amount of environmental data has increased considerably over the last decade - so much that soil scientists could become overwhelmed by the “data deluge”, the “data revolution” represents a positive challenge for soil science. But to achieve this goal, it is crucial for soil scientists to be involved alongside data and computer scientists to ensure that expert knowledge is captured and used to validate results, and to acknowledge and preserve legacy data.

The fourth paradigm is a new way of generating scientific knowledge building on the “data deluge” and on the “big data” technologies. Associated upcoming changes require scientists to adapt their tools and re-skill themselves to fully benefit from these developments. In particular, a shift towards data-driven approaches requires data to be openly available, support reproducibility of the scientific results, and foster the upscale of those results. Ensuring the interoperability and availability of soil data is not only a requirement for the emerging sensing technologies, but can also improve the visibility of the staggering amount of soil information that is currently inaccessible. In particular, setting up soil data services is an import step to improve the dissemination of soil information, and support its use by very various users and across a range of sectors outside of the sole research field of soil science.

References

- [1] Hilbert M 2013 Big data for development: from information to knowledge societies *SSRN Electronic Journal* <http://ssrn.com/abstract=2205145>
- [2] Hilbert M and Lopez P 2011 The World’s technological capacity to store, communicate, and compute information *Science* **332** 60-65
- [3] Gray J 2009 Jim Gray on eScience: a transformed scientific method *The Fourth Paradigm: Data-Intensive Scientific Discovery* ed Hey T, Tansley S and Tolle K (Redmond, WA, USA: Microsoft Research)
- [4] Gartner Inc 2011 Gartner says solving 'Big data' challenge involves more than just managing volumes of data <http://www.gartner.com/newsroom/id/1731916>
- [5] Cukier K 2010 Data, data everywhere: a special report on managing information *The Economist* **394** 3-5
- [6] Boyd D and Crawford K 2012 *Information, Communication & Society* **15** 662-79
- [7] Lagacherie P and McBratney AB 2006 Spatial soil information systems and spatial soil inference systems: Perspectives for digital soil mapping *Developments in Soil Science Volume 31 Digital Soil Mapping An Introductory Perspective* ed P Lagacherie AB McBratney and Voltz M (Elsevier) pp 3-22
- [8] Adamchuk VI and Viscarra Rossel RA 2010 Development of on-the-go proximal soil sensor systems Proximal Soil Sensing *Progress in Soil Science* ed Viscarra Rossel R A, McBratney AB and Minasny B (Springer Netherlands) pp 15-28
- [9] Mulder VL, de Bruin S, Schaepman ME and Mayr TR 2011 The use of remote sensing in soil

- and terrain mapping – A review *Geoderma* **162** 1-19
- [10] Hey T, Tansley S and Tolle K 2009 *The Fourth Paradigm: Data-Intensive Scientific Discovery* (Redmond, WA, USA: Microsoft Research)
 - [11] Bouma J 1989 Using soil survey data for quantitative land evaluation *Advances in Soil Science* (*Advances in Soil Science* no 9) ed Stewart B A (Springer US) pp 177-213
 - [12] de Gruijter J, Brus DJ, Bierkens MFP and Knotters M 2006 *Sampling for Natural Resource Monitoring* (Springer)
 - [13] McBratney AB, Mendona Santos ML and Minasny B 2003 On digital soil mapping *Geoderma* **117** 3-52
 - [14] Behrens T and Scholten T 2006 A comparison of data-mining techniques in predictive soil mapping *Developments in Soil Science Volume 31 Digital Soil Mapping An Introductory Perspective* ed P Lagacherie A B M and Voltz M (Elsevier) pp 353-617
 - [15] Bui EN and Moran CJ 2001 Disaggregation of polygons of surficial geology and soil maps using spatial modelling and legacy data *Geoderma* **103** 79-94
 - [16] Bui EN, Loughhead A and Corner R 1999 Extracting soil-landscape rules from previous soil surveys *Aust J Soil Res* **37** 495-508
 - [17] Bui EN, Henderson BL and Viergever K 2006 Knowledge discovery from models of soil properties developed through data mining *Ecol Modell* **191** 431-46
 - [18] Ibanez L, Schroeder WJ and Hanwell MD 2014 Practicing open science *Implementing reproducible research* ed Stodden V, Leisch F and Peng RD (Florida: CRC Press) 241-280
 - [19] Wolkovich EM, Regetz J and O'Connor MI 2012 Advances in global change research require open science by individual researchers *Global Change Biology* **18** 2102-10
 - [20] Ince DC, Hatton L and Graham-Cumming J 2012 The case for open computer programs *Nature* **482** 485-88
 - [21] Malik T, Pham Q and Foster I 2014 SOLE: towards descriptive and interactive publications *Implementing reproducible research* ed Stodden V, Leisch F and Peng RD Peng (Florida: CRC Press)
 - [22] Stodden V, Guo P and Ma Z 2013 *PLoS ONE* **8**
 - [23] Russel J 2013 Reproducing our irreproducibility *Nature* **496** 398-98
 - [24] Bloom T 2014 Data access for the open access literature: PLOSs data policy <http://www.plos.org/data-access-for-the-open-access-literature-ploss-data-policy/>
 - [25] Smith A 2014 Improving GitHub for science <https://github.com/blog/1840-improving-github-for-science>
 - [26] Heidorn PB 2008 Shedding light on the dark data in the long trail of science *Libr Trends* **57** 280-99 ISSN 1559-0682
 - [27] Rossiter DG 2008 Digital soil mapping as a component of data renewal for areas with sparse soil data infrastructures *Digital Soil Mapping with Limited Data* ed Hartemink AE, McBratney A and Mendona- Santos MdL (Springer Netherlands) pp 69-80
 - [28] Hempel J, McBratney A, Arrouays D, McKenzie N and Hartemink A 2014 *GlobalSoilMap project history GlobalSoilMap: Basis of the global spatial soil information system* ed Arrouays D, McKenzie N, Hempel J, Richer de Forges A and McBratney AB pp 3-8
 - [29] Keune H, Murray AB and Benking H 1991 Harmonization of environmental measurement *GeoJournal* **23** 249-55
 - [30] Arrouays D, Marchant BP, Saby NPA, Meersmans J, Orton TG, Martin MP, Bellamy PH, Lark RM and Kibblewhite M 2012 Generic issues on broad-scale soil monitoring schemes: A review *Pedosphere* **22** 456-469
 - [31] Baume O, Skien J O, Heuvelink G B M, Pebesma E J and Melles S J 2011 *Inter J App Earth Obs* **13** 409-419
 - [32] Bizer C, Heath T and Berners-Lee T 2009 *Int J Semant Web Inf* **5** 1-22
 - [33] Simons B, Wilson P, Ritchie A and Cox S 2013 ANZSoilML: an Australian-New Zealand standard for exchange of soil data *EGU General Assembly Conference Abstracts* **15** 6802

- [34] Percivall G 2010 *Int J Digital Earth* **3** 14-30
- [35] Müller M, Medyckyj-Scott D, Cowie A, Heuer TH and Roudier P 2013 Current status and future directions of mobile GIS ed Moore A and Whigham PA (University of Otago, Dunedin, New Zealand)
- [36] The James Hutton Institute SOCiT URL <https://itunes.apple.com/gb/app/socit/id631266307>
- [37] Goodchild MF 2007 Citizens as sensors: The world of volunteered geography *GeoJournal* **69** 211-221
- [38] Jacobs A 2009 *Commun ACM* **52** 36