The International Journal of Digital Curation
Issue 1, Volume 6 | 2011

The Milieu and the MESSAGE: Talking to Researchers about Data Curation Issues in a Large and Diverse e-Science Project

Martin Donnelly,

Digital Curation Centre,

University of Edinburgh

Robin North,

Centre for Transport Studies,

Imperial College London

Abstract

MESSAGE (Mobile Environmental Sensing System Across Grid Environments) was an ambitious, multi-partner, interdisciplinary e-Science research project, jointly funded by the Engineering and Physical Sciences Research Council (EPSRC) and the UK Department for Transport (DfT) between 2006 and 2009. It aimed to develop and demonstrate the potential of diverse, low cost sensors to provide heterogeneous data for the planning, management and control of the environmental impacts of transport activity at urban, regional and national level. During the last year of the project, the Digital Curation Centre (DCC) interviewed and observed members of the project team in order to identify and analyse key aspects of their data-related activities, recording attitudes towards the data that they create and/or re-use. This paper describes the major issues identified over the course of the case study, which are presented in parallel with the perspectives of the project team in order to demonstrate the multiplicity of views that may be projected onto a single dataset. It concludes with a contextualisation of the case study's themes with those of a number of contemporary reports.[1]

[1] This paper is based on the paper given by the authors at the 6th International Digital Curation Conference, December 2010; received December 2010, published March 2011.

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. ISSN: 1746-8256 The IJDC is published by UKOLN at the University of Bath and is a publication of the Digital Curation Centre.

Introduction

MESSAGE (Mobile Environmental Sensing System Across Grid Environments) was a three-year research project, running from October 2006 to September 2009, funded jointly by the Engineering and Physical Sciences Research Council and the UK Government's Department for Transport (DfT), and comprising a multi-partner consortium: academic research groups (the Universities of Cambridge, Leeds, Newcastle and Southampton, and the project's lead institution, Imperial College London); private sector companies; and local and national government agencies.

During the last year of the project, a member of the UK's Digital Curation Centre (DCC) was allowed to sit in on the project meetings and interview key personnel in order to prepare a case study on the data-related aspects of the work, examining attitudes towards data from a range of roles within the project consortium: research scientists, data managers, and project coordinators.

Established in 2004, the DCC is the UK's leading centre of expertise in digital data curation. The Centre's mission is partly evangelical: to advocate the adoption and ongoing dissemination of good practice within and beyond the UK Higher Education sector. In the DCC's second phase (2007-2010), the strategy was to engage directly with researchers, while in its current, third phase the strategy has shifted towards targeting key intermediaries. As an item of work that spans the two phases, the DCC-MESSAGE study is therefore situated at an interesting juncture, allowing the DCC to learn about researchers' individual perspectives on curation issues, as well as how they feel these relate to their ongoing dealings with, for example, their funding bodies.

This paper summarises the findings of the case study, the principal theme of which is that datasets can only be adequately considered from a variety of perspectives, and provides insight into the attitudinal and practical differences which existed among the various stakeholders from both an internal and an external standpoint. We conclude by giving a contextual cross-comparison between our findings and the themes identified in a number of relevant contemporary reports.

MESSAGE

Pollution and Sensors

Urban air pollution, including sulphur dioxide (SO_2), ozone (O_3) and nitrogen oxides (NO_x), is a major health problem around the world (Air Quality Expert Group, 2004), with estimates suggesting that traffic and transport activity is the dominant source in many cities (European Environment Agency, 2007). Appropriate monitoring of the atmospheric conditions, combined with interventions in traffic management and the provision of traveller information can help to mitigate these problems (Tate & Bell, 2000). The MESSAGE system was conceived to explore the opportunities provided by a new generation of low-cost and mobile sensor systems, and their potential for integration within an intelligent sensing network. In this, MESSAGE built on concurrent advances in communications, positioning, computing, sensing and modelling technologies to implement a mobile, wireless environmental sensing network and data processing infrastructure. With this network, significantly richer, and larger, datasets about the dynamics of urban air pollution can be captured. The effective management and curation of these data, in a politically sensitive arena with multiple stakeholder agencies, is therefore a considerable challenge.

Antecedents

MESSAGE found its roots partially in an earlier project: the National Transport Data Framework (NTDF). This was also funded by the DfT and was a collaboration between the Cambridge e-Science Centre and the Centre for Transport Studies at Imperial College London. The NTDF collected various transport-related data resources, including traffic flows, events on the transport network (such as accidents, scheduled engineering work), railway timetables, and put together a search engine which enabled people to discover these sorts of data more readily. The data were annotated to work in conjunction with ontologies, making searches more intelligent and enabling users to discover sources of data based on keywords. MESSAGE grew in part out of that collaboration, with the goal of building an infrastructure to collect a different sort of data, namely pollution data, at much more fine-grained time and spatial scales than had previously been done.

Aims and Objectives

MESSAGE had two inter-linked aims: (i) to investigate, develop, harness and demonstrate the potential of diverse, low-cost and semi-ubiquitous sensors to provide data for the planning, management and control of the environmental impacts of transport activity at urban, regional and national levels, and reporting pollution levels to end users in real-time; and (ii) to develop a flexible, scalable and reusable Grid-based e-Science infrastructure for transferring and processing the data captured by this heterogeneous sensor network, and supporting a wide range of scientific, policy-related and commercial applications. This infrastructure involved linking middleware applications to process incoming data (both from the MESSAGE sensors themselves, deployed by application scientists, and from third-party data sources) to provide greater value to potential users of the system.

Data in the MESSAGE Project

The project team demonstrated how low-cost sensors – both mobile and static – could be deployed in high densities and linked wirelessly, sharing information and thereby minimising data logging and data handling issues. (Figure 1.) The MESSAGE datasets are primarily observational, i.e. non-repeatable, and therefore make strong candidates for preservation and curation.

The integration of sensors with mobile devices that people generally carry with them most of the time (i.e. mobile phones) with vehicles and street infrastructure effectively creates a pervasive network of roaming sensors. This lays the groundwork for the creation of a larger archive dataset of pollution levels which could act as input to pollution models to help predict more accurately what pollution levels might be like based on their current values, local weather conditions and other factors, and thereby provide an improved basis for the evaluation and implementation of policies to improve local air quality.

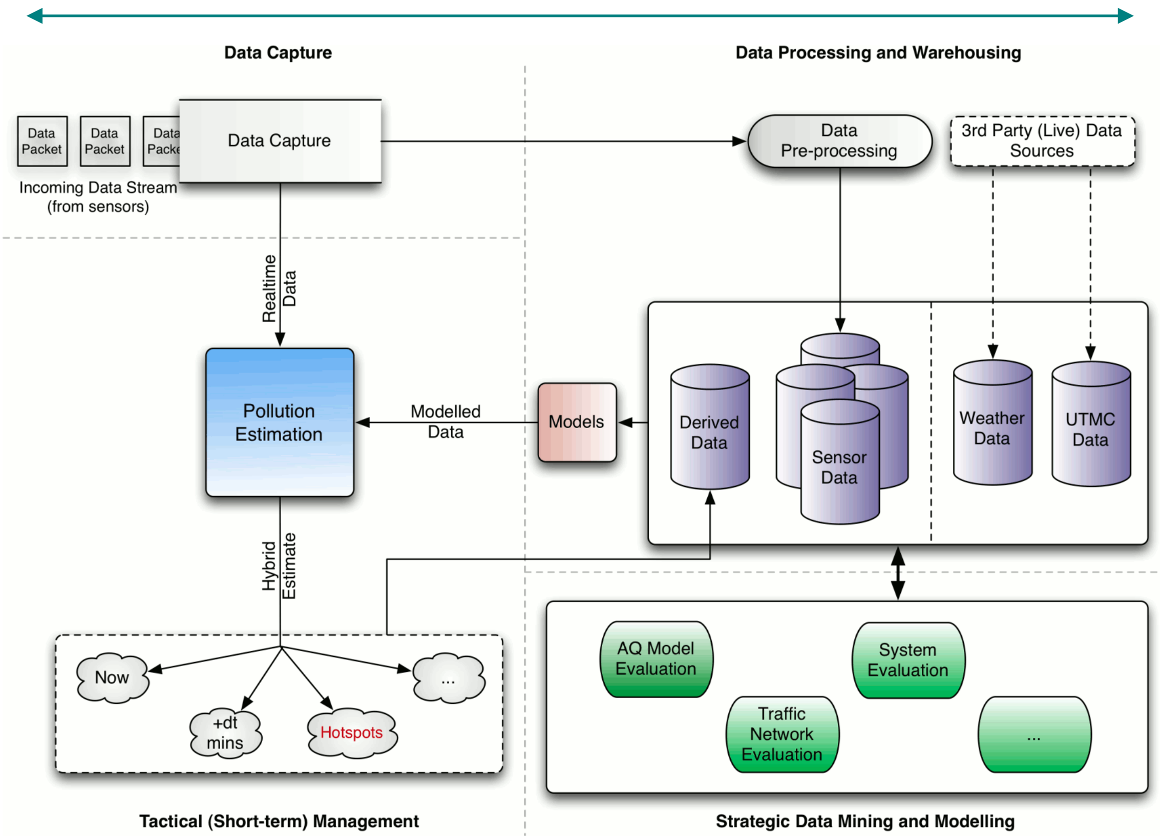

Figure 1. Data Modelling, Mining and Evaluation to Support Environmental Management (Cohen, North et al., 2009)

Sensor Development

Each of the academic partners had its own particular focus, with three of the universities developing distinctive sensor platforms: Cambridge investigated the potential for personal devices (mobile phones) to support a sensing system; Newcastle developed a “smart-dust” network using ZigBee (IEEE 802.15.4) motes for mounting on street furniture; and Imperial devised a network utilising Wi-Fi (IEEE 802.11g) and WiMAX (IEEE 802.16) technologies for communications and positioning, together with a set of novel sensor designs for use on vehicles and as high-resolution reference units for the other sensors. Southampton’s main involvement was in the area of evaluation and user studies, while Leeds’ role was to quality assure the data captured by the experimental mobile sensors against their own well-established, static sensor technology.

The mobile sensors were variously mounted on vehicles or carried by walking and cycling humans in order to act as mobile, real-time environmental probes, sensing transport and non-transport related pollutants and hazards. Each of the sensor platforms was designed to integrate with the e-Science architecture, with differences in data structure reflecting the differing processing and communications capabilities of the sensor hardware.

Within the sensor strand of the project there were two sub-areas of concentration, exemplified here via sample uses:

> **Longer-term data to support strategic objectives.** Transport researchers, local authorities and transport companies may wish to examine emissions from traffic over periods of time, and adjust their policies, strategies or models either to study, mitigate or minimise emissions, or to inform changes in routing;
>
> **Real-time data to support operational decision-making.** These data may be used by network operators to refine their system management, or by individuals to inform their own travel choices. For example, someone with an asthmatic or breathing condition may wish to see a local pollution forecast before travelling, and determine which route to their destination exposes them to the least pollution over the journey's course.

e-Science Infrastructure

The other primary aim of the project was to develop a flexible e-Science infrastructure to support the heterogeneous sensor deployments. This strand of work was led by the Internet Centre within the Department of Computing at Imperial College and involved broad collaboration across the project partners, and with external organisations.

The MESSAGE e-Science infrastructure handles the data captured by the heterogeneous sensor deployments across multiple (and sometimes mobile) sites. Data imported from the various sensors carries similar properties: all of the sensors record time, location/position and readings for various pollutant species – the mobile sensors also record their speed and bearing – and other, quality-related properties such as noise and calibration settings may be recorded.

This then enters the data management infrastructure, the layer between the sensors and the databases (Figure 2). Pre-processing is carried out to transform the data from the different native capture formats used for simplified wireless data transmission from the sensor nodes into a standard XML transport format, enabling interoperability and movement of data within a distributed service-oriented architecture. Early in the project it was decided that each of the partners would follow their own data capture and processing workflow, with a common data storage schema based on the Urban Traffic Management and Control (UTMC) specification[2]. This enabled the data management infrastructure to deal with differences in capture devices and their captured data, encourage efficiencies, and support the ready integration of future sensor systems. An additional benefit of this process was the ability to test multiple data capture approaches prior to determining which was the most efficient for a larger scale future deployment, with potentially millions of sensors online at once nationally, or even internationally. A scalable data handling system was developed at Imperial using web services and comprising multiple gateways to accept data from an arbitrary number of sensors and pass it into the central management area.

[2] Urban Traffic Management & Control: http://www.utmc.uk.com/.

Gateways can only handle a limited amount of simultaneous information input, so the numbers of gateways required at any given time is a function of the number of sensors feeding into the system and the frequency with which they are reporting. This has the potential to vary dramatically over the course of a day, from large numbers of mobile sensors online during, for example the morning rush hour, to fewer sensors in the middle of the night.[3] Researchers investigated emerging cloud computing technologies as a method to provide on-demand, scalable computing infrastructure to meet the need for significantly varying computational requirements, and to minimise unused resource and the likelihood of overload.

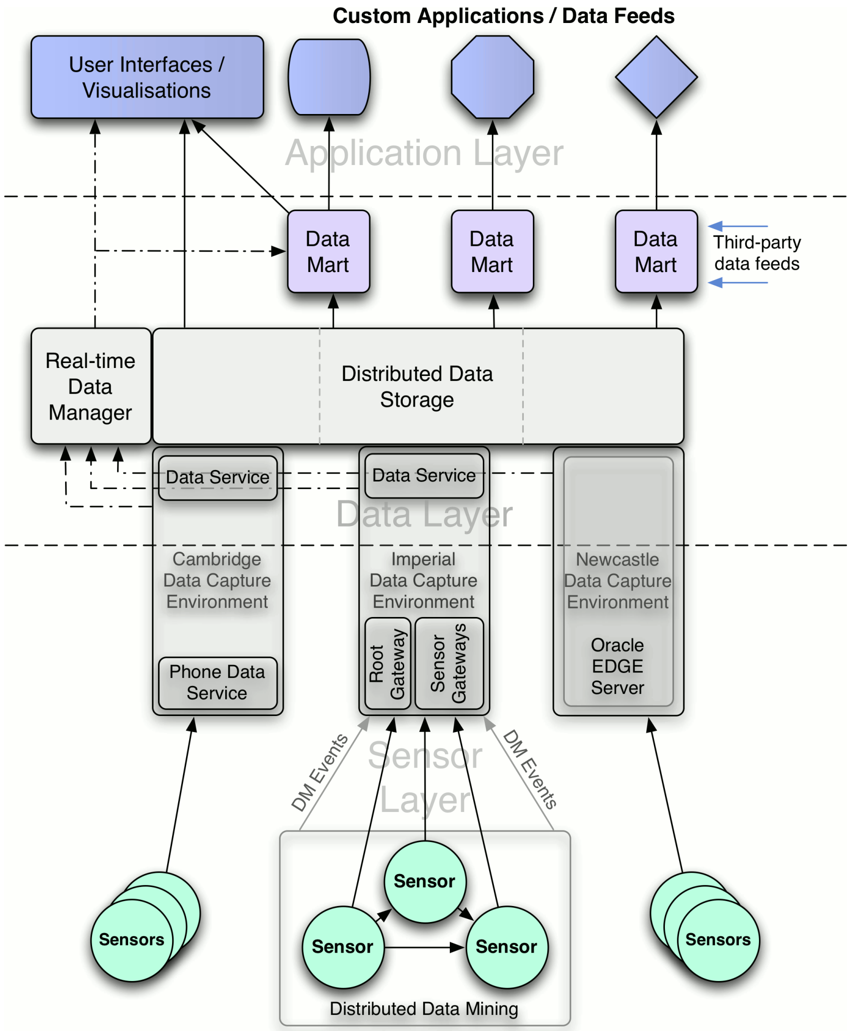

Figure 2. The MESSAGE e-Science Architecture (Cohen, 2009)

[3] An international roll-out would, of course, reduce to a certain extent these peaks in demand.

The MESSAGE-DCC Collaboration

Context

In its second phase, the Digital Curation Centre (DCC) included an e-Science Liaison strand in its work plan. One of the major outputs of this was the publication of Graham Pryor's longitudinal study of the CARMEN project (Pryor, 2008). Undertaken between December 2008 and November 2009, the engagement between the DCC and MESSAGE represented the second of these longitudinal studies. It tracked the emergence of technological and organisational solutions to a range of discrete and heterogeneous data-handling problems spanning a variety of domains, including Transport Studies, Computer Science, Environmental Science and Engineering.[4]

Methodology

The primary source for the study was a series of nine interviews carried out with key members of the MESSAGE team, spanning researchers, evaluators, dedicated computing staff and project coordinators. Interviews were semi-structured and most were carried out in person, usually at the subject's home institution. On the few occasions when this was not practical, interviews were carried out via telephone. The interviews were recorded and transcripts were provided to the subjects in order to give them an opportunity to correct any errors.

In addition to the interviews, the DCC reporter was granted access to the project's internal wiki and was invited to attend (as a silent observer) several meetings of the project's Scientific Committee, and a major public "demonstration day" towards the end of the project.

Varying Perceptions of Data-Related Issues

While the MESSAGE project did not in itself capture large amounts of data, it lays the technical and infrastructural groundwork for much larger scale future data collection from heterogeneous sources and sensor types. Among the DCC's driving concerns was that huge datasets may be created in the near future without adequate planning – operational or financial – for their ongoing management and curation: in other words, as mobile sensor technologies become more accurate and unit costs fall, the volume of data captured will balloon. Yet, at the same time, these improvements coincide with a global situation where there are fewer resources available (financial, and therefore human) to manage it appropriately.

From the beginning of the MESSAGE project, there was an emphasis not just on developing mechanisms to collect data, but also to ensure that these data were managed and integrated in such a way as to support new insights into the nature of transport and urban air pollution, and consequently the development of more effective scenario modelling and decision support tools. These objectives were not developed with knowledge of the DCC's Curation Lifecycle Model,[5] and were framed more in terms of data management than curation. The focus was primarily on the contemporary mechanisms for processing and managing the data, and consideration of the longer-term value of the data management *structure*, and in particular its generic application

[4] Other touch points included the suite of SCARP case studies (http://www.dcc.ac.uk/projects/scarp), and the RIN-funded Case Studies in Life Sciences led by the Institute for the Study of Science, Technology and Innovation (ISSTI) and the DCC (2009).
[5] DCC Curation Lifecycle Model: http://www.dcc.ac.uk/resources/curation-lifecycle-model.

to other areas, was therefore a separate consideration beyond the value of the data collected during the project itself.

Commentary on Emerging Themes

The following table lists a subset of the themes of the case study report (i.e. those pertinent to this paper's topic), together with a brief extrapolation of the theme from an outsider's standpoint and the internal view drawn from interviews with members of the project team.[6]

The findings are presented side-by-side in order to facilitate comparison of the viewpoints; the final case study report (Donnelly, 2010) itself takes differences in standpoint and perception as its principal motif.

Theme	*DCC perception*	*MESSAGE perception*
Ways of looking at a dataset (human-centeredness and domain-specificity of curation)	Different cohorts brought different expectations and goals to the project as a whole, just as different communities have different ideas about what a given dataset will do for them. Some researchers will want to "do science" on the data as soon as possible, while others may have different intentions altogether: the data may be just another resource in their work. Caught in the middle is the data manager. As one interviewee noted: "It is very, very tricky to work interdisciplinarily and inter-institution where everyone's requirements and needs are always different. There is always something, as a data manager, you need to accomplish which is always hidden behind the agenda of others."	With a new system and a new type of data, it is important to ensure that appropriate quality assurance procedures are implemented before people attempt to use the data for its intended application. During this process the data will be evaluated in different ways to those familiar to the application domain experts and that can cause tension. Another interviewee expressed the view that "it is always difficult to conceive of how data will be used until you actually come to try and use it." This obviously impacts on the creation of data management plans, as a later theme shows.
The time-sensitivity of data	The nature of a dataset will change inevitably over time from different perspectives as its context, position and importance shifts, even if the data itself does not change. Perceptions of the role and position of the datasets did indeed change over the course of this project.	One non-domain scientist noted his original view, that the real-time data would be the most important and the most valuable resource or facet of the system, had, in time, been joined by the perception that the historical analysis of large quantities of older, legacy data was actually just as important, and perhaps even more so.

[6] N.B. The themes were identified externally, by the DCC observer.

Theme	DCC perception	MESSAGE perception
Security and sensitivity of data: barriers to data sharing	Much of the data collected by the MESSAGE sensors seemed from an external perspective (and certain internal perspectives) to be fairly mundane, harmless stuff, but after a little probing it became clear that the potential for controversy existed within it in latent form. Even simulated "non-data" had the potential to cause embarrassment if not handled and labelled correctly. One interviewee summarised the issue succinctly: "The degree of sensitivity is not to the data itself, it's to the interpretations that could be put on the data…" Barriers to data sharing may therefore emerge from many sources: there are self-imposed barriers (relating to confidence or lack thereof in the quality of data and documentation); internal influences such as research group agreements and institutional ethics policies; external influences such as legal and regulatory issues; and perhaps even technical barriers, although these appear to be the easiest to overcome.	From the project team's perspective, issues relating to data availability and management were satisfactorily addressed, although some issues were encountered with partners for whom the data were felt to be politically sensitive. This led to the implementation of more elaborate data management procedures and authentication processes than had initially been envisioned. Particular sensitivities surrounded the potential for short-term peak pollution levels that were revealed by the new sensing techniques (which may not affect overall averages) to be inappropriately disseminated, leading to public disquiet and political embarrassment. This became an issue in particular where measurements might be misinterpreted when compared to prevailing regulatory measurements (i.e. similar units, but different sampling periods), and several external partners placed firm restrictions on data availability.
Data trade within the research field, and beyond it	Many researchers are warily supportive of increasingly open access to their data, but that they have concerns about the standard to which their data is documented, and therefore about how useful it would be to third party users. Similarly, researchers tend to take a passive approach to data sharing. In the main they are happy for others to use their data, but they seldom go out of their way to make this happen by making their data easy to find, for example, by depositing it in a repository. This underlines the view that expecting researchers to manage data for future use without reward (or even funding) is an unsustainable strategy – it will always play second fiddle to the tasks that gain reward and recognition.	Within the project, efforts were made to share data between institutions to test for the transferability of the data types and the meta-data descriptors. This led to refinements in the data specification process. Long term, open access to data requires pro-active maintenance of a repository and an openness to answering future queries in relation to the datasets. This appears hard to achieve under current funding structures.
Agreements: formal and informal (formalisation of data-related agreements vs perception of additional bureaucracy)	It is reasonable to infer that a lack of precise allocation of responsibility for data management for the longer term leads to a risk that the datasets will be lost, or become inaccessible or incomprehensible, which may be worse. When asked who was ultimately responsible for the data over the longer term, one interviewee laughed and said: "I suppose I assume it's not me." A culture of assumptions and lack of clearly defined responsibility unquestionably leads to risk, which can be managed to a considerable degree by the development and adoption of data management strategies early in a project's lifecycle.	One senior interviewee warned of the danger of creating "additional bureaucracy that only slows down the creation and starting of projects," questioning whether the existing arrangements were really inadequate, and suggesting that the creation of "model" agreements might help solve the issue in an efficient way. Another interviewee noted that the transient nature of academic research posts mitigates against this, as it is often difficult to document working arrangements sufficiently well so as to allow future researchers to collaborate on the same basis.

Theme	DCC perception	MESSAGE perception
Data Management Plans (and the importance of planning)	At the grant application and post-funding stage, there was no pressure from the project's funders to come up with a data management strategy or plan, but when asked, some interviewees acknowledged that data management is an area of greater complexity than they had initially considered, and that as the project (and case study) developed, overall awareness had increased. The final case study report recommends that research funders should support researchers to develop and maintain data management plans (and other formalised agreements) from the grant application stage onwards, and that data sharing policies should be emphasised and project compliance monitored.	Late in the project, more than one interviewee expressed the view that the idea of paying more attention to data curation concepts earlier in the project lifecycle has to be right, and noted that through the case study process they had come around to the idea that time spent considering data management issues early in the project preparation stage more than pays for itself in the longer term. However, there was also an emphasis on the need for external support and training in doing this; from the funder, or from national bodies such as the DCC. In a climate of cuts and "make-do-and-mend", demand for data reuse is only going to increase, and the implications for supporting this at scale must be addressed now.
The 'keep everything' mentality	When asked about appraisal and selection methods, all of the researchers claimed to keep everything that they created, with one adding that "no data is ever obsolete." From an external standpoint, this is counterintuitive: the more data held, the more difficult it becomes to find what you are looking for. If, as the saying goes, the best place to hide a leaf is in a forest, the worst place to look for a particular datum is inside a computer crammed full of data.	In research projects where substantial system prototyping is involved, data tends to be archived in blocks relating to particular phases of the system development. Often this is kept to facilitate future debugging of as yet unknown problems. This tends to teach us to keep everything so that in the very worst case, the raw data can be reconstructed.

A related theme not addressed by the case study but pertinent to this paper is that of training. More widespread and embedded training may have impacted positively on some of the above themes, especially data planning and appraisal: one interviewee noted that "the nature of research projects is that very often the personnel involved have not worked on anything similar before. This may be especially true when looking at large inter-disciplinary projects." Similarly, early-career researchers tend to be quite mobile and expertise will often move on at the end of funded projects.

Training is often cited as a cornerstone of improved future curation, but when "learning by doing" appears to be so prevalent within the academy, and academics often seem determined to create their own systems and methodologies, the question which follows is: where does this leave the DCC and other training bodies?

Conclusions

The study's core message is that objectivity towards data is neither constructive nor truly possible, and that data's position within the larger research system (or system of systems) changes inevitably with the passage of time. The major recommendation is therefore that research funders and support organisations should continue to endeavour to find perspective-driven means of easing the data management process via the development (and potential mandating) of policies, guidance and tools with a user focus in mind.

Other themes which emerged during the case study support and relate to the findings of other recent reports, and a brief synthesis is presented here. The key conclusion of RIN (2009) is that the policies and strategies of research councils and information service providers must be informed by an understanding of the exigencies and practices of research communities if they are to be effective in optimising the use and exchange of information, and in ensuring that this is scientifically productive and cost-effective. The authors note that a single approach to the future of life sciences or a one-size-fits all information policy will not be productive or effective, and that information service policies and provision need to be brought closer to research groups and communities.[7]

On the merits of data management planning, Van den Eynden et al. (2010) found that researchers from completed projects reported the creation of data management plans to be beneficial, but that that they needed "clear information on how to plan data management in a meaningful way, and often need additional support to develop good management procedures." They also warn that "[p]lanning data management does not guarantee its implementation, and research funders need to consider how to ensure that good data management intentions are indeed implemented and revisited."

On incentivisation/motivation, as Lyon (2009) notes, UK research assessment and most journal publishers do not reward data sharing, social web contributions or peer production approaches to data curation, and that this lack of incentive for data sharing and participatory methodologies acts as a barrier to data sharing and reuse. Similarly, the RIN authors note that grant-holders will "do what is necessary to be seen to fulfil [funders'] requirements", but that practical and human issues – including lack of formalisation, clarity of description, planning, recognition and reward – serve to restrict exchange. The key element in this quote is: "be seen to fulfil". The opening of data stands at risk of being reduced to an obligation: a box-ticking exercise.

On funding, the RIN authors note the difficulties involved with sustaining data management roles and resources in the fixed-term project-based funding model, particularly in smaller groups. Lyon also notes that the economic implications of data sharing will require detailed analysis, and the societal benefits will need to be reviewed and evaluated as part of a larger exercise.

On the provision of tools and guidance, the RIN authors found that "researchers are reluctant to adopt new tools and services unless they know a colleague who can recommend or share knowledge about them." They also note the importance of tangible results in tool uptake and the necessity of fitting in with existing research workflows and schedules. Van den Eynden et al. found that researchers want "easy, practical and trustworthy solutions they can embed into research activities, rather than a range of guidelines or suggestions from which to choose." This chimes with the recommendation of RIN that "better (i.e. easy-to-use) tool-based support be provided to researchers to enable them to undertake their own data curation."[8]

In an attempt at growing the curation community, engagement with researchers must be pitched at an appropriate level (domain by domain, if not discipline by discipline), and couched in language that they already speak: the RIN authors urge funders to adopt a more pragmatic and experimental policy from the bottom-up,

---

[7] In furtherance of this message, the RIN has funded a further two sets of cases studies in this series, on the Humanities and the Physical Sciences.

[8] Tools given as exemplars by the RIN authors are the DCC's Data Audit Framework, DRAMBORA, and Data Management Planning resources such as DMP Online.

recognising the multiplicity of contexts and different approaches to information sharing, and building on the informal sharing that is already taking place, based on the recognition of mutual needs.

To conclude, the authors agreed – from their respective standpoints – that human data-related issues are considerably more difficult than technological data-related issues, and the case study observations support the view expressed widely in recent years that increased effort must be made to embed sound data management practice from the outset in order to ensure longer-term access to the data that underpins the records of science.

Acknowledgements

This case study on which this paper is partly based is an output of the DCC's Phase 2 work, funded by the Joint Information Systems Committee (JISC), to whom thanks are due. On the MESSAGE side, particular thanks to the funders and project sponsors – especially the EPSRC and the Department for Transport – and the researchers and data managers who were interviewed for the case study. More information about the project and is available at www.message-project.org.

Disclaimer

The views expressed in this paper are those of the authors, and do not represent the view of the Department for Transport or any of the non-academic partners of the MESSAGE project.

References

Air Quality Expert Group (2004). *Nitrogen dioxide in the UK.* Department of Environment Food and Rural Affairs (DEFRA).

Cohen, J. (2009). Mobile environment sensing system across grid environments: e-Science architecture and computing infrastructure. *Presented at the Combined MESSAGE Project Demonstration Day.* London.

Cohen, J., North, R., et. al. (2009) Creating the MESSAGE Infrastructure. *Traffic Engineering and Control.* London: Hemming Information Services.

Donnelly, M. (2010). *Ways of looking at data: MESSAGE (Mobile Environmental Sensing Mobile Environmental Sensing System Across Grid Environments).* Digital Curation Centre (DCC). Retrieved December 29, 2010, from http://www.dcc.ac.uk/resources/case-studies/message.

European Environment Agency (2007). *Europe's environment – The fourth assessment.* EEA. ISBN: 978-92-9167-932-4.

Lyon, L. (2009). *Open science at web-scale: Optimising participation and predictive potential.* Consultative report for the Joint Information Systems Committee (JISC). Retrieved December 29, 2010, from http://www.jisc.ac.uk/publications/reports/2009/opensciencerpt.aspx.

Pryor, G. (2008). *CARMEN: Code, analysis, repository & modelling for e-Neuroscience.* Digital Curation Centre (DCC). Retrieved December 29, 2010, from http://www.dcc.ac.uk/sites/default/files/documents/publications/case-studies/carmen.pdf.

Research Information Network (2009). *Patterns of information use and exchange: Case studies of researchers in the life sciences.* RIN. Retrieved December 29, 2010, from http://www.rin.ac.uk/system/files/attachments/Patterns_information_use-REPORT_Nov09.pdf.

Tate, J.E., Bell, M.C. (2000). Evaluation of a traffic demand management strategy to improve air quality in urban areas. *Proceedings of the 10th International Conference on Road Transport Information and Control.* IEEE, Conf. Publ. No. 472.

Van den Eynden, V. et al. (2010). *Data management practices in the social sciences.* UK Data Archive (UKDA), Retrieved December 29, 2010, from http://www.data-archive.ac.uk/media/203597/datamanagement_socialsciences.pdf.