

The International Journal of Digital Curation

Issue 2, Volume 6 | 2011

Curating Complex, Dynamic and Distributed Data: Telehealth as a Laboratory for Strategy

Jenny Ure and Janet Hanley,
Edinburgh Napier University

Tasneem Irshad, Angus Whyte, Claudia Pagliari, Hilary Pinnock and Brian McKinstry,
Edinburgh University

Abstract

Telehealth monitoring data is now being collected across large populations of patients with chronic diseases such as stroke, hypertension, COPD and dementia. These large, complex and heterogeneous datasets, including distributed sensor and mobile datasets, present real opportunities for knowledge discovery and re-use, however they also generate new challenges for curation. This paper uses qualitative research with stakeholders in two nationally-funded telehealth projects to outline the perceptions, practices and preferences of different stakeholders with regard to data curation. Telehealth provides a living laboratory for the very different challenges implicit in designing and managing data infrastructure for embedded and ubiquitous computing. Here, technical and human agents are distributed, and interaction and state change is a central component of design, rather than an inconvenient challenge to it. The authors argue that there are lessons to be learned from other domains where data infrastructure has been radically rethought to address these challenges.

Context

Data collected through sensors, mobile phones and telemetrically-supported devices in the home now facilitate remote monitoring of chronic conditions such as hypertension and chronic obstructive pulmonary disease (COPD). Multi-dimensional data, collected from multiple individuals over time, provides a unique new resource for research, as well as for more individualized diagnosis and treatment. Dealing with data of this complexity and scale challenges existing practice at every stage of data management. In this rapidly evolving healthcare landscape it is not clear what the roles and responsibilities are for curating and storing these datasets, and what resources and incentives would be available to achieve this. In many senses, practice has outstripped the scope of existing policy and infrastructure for governance, as well as for curation, analysis and use. In the interim, this data is not being managed in ways which would allow the many fragmented telehealth research studies to easily share or re-use these resources. The cost of generating these datasets, the scale and speed with which they are being used, and the potential value for research makes this an urgent challenge for data curation if these resources are to provide a return on investment in future. The paper draws on research carried out on data curation challenges in telehealth monitoring by the authors (Irshad & Ure, [2009](#)) as part of the Digital Curation Centre SCARP Study project on data curation practices in different disciplines (Lyon et al., [2010](#)).

The telehealth projects that were the focus of the research were pilot studies evaluating the use of home-based telemetry to measure symptoms and vital parameters such as blood pressure (BP), blood glucose and pulse-oximetry using linked monitoring devices (McKinstry et al., [2009](#)). Tele-monitoring devices provide a system for prompting patients to take their medications and record their symptoms (aided by bluetooth connectivity to monitoring devices), with the potential for voice or video-consultation as well. Data submitted to the system are transmitted (in most cases) to a central call centre manned by trained support staff, who may contact the patient or their health care providers if readings are out of range. Figure 1 gives an example of the typical lifecycle of tele-monitoring data.

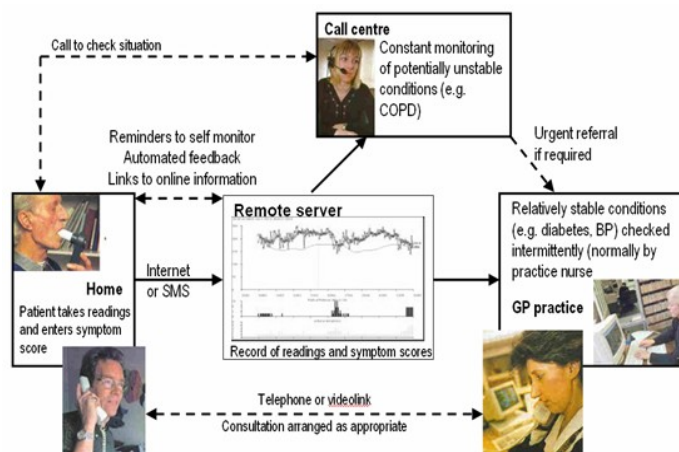


Figure 1. From home to healthcare centre - the lifecycle of telehealth data. (Image courtesy of Dr Hilary Pinnock)



As the figure indicates, data passes through many hands in transit from home to healthcare centre, with many interfaces between people, processes and technologies. The aim of the two target studies we drew upon was to evaluate the potential of tele-monitoring systems at home to reduce hospital admission through early intervention in COPD (COPD study) and hypertension (HITS study). This provided a laboratory for strategy in implementation, but also for consideration of the issues in data collection, management, curation and re-use that would be necessary in large scale implementation.

Research Methods

A qualitative approach (Denzin & Lincoln, [2000](#); Strauss & Corbin, [1990](#)) was adopted with transcription and independent coding of interviews and focus groups by two researchers, from which a set of evolving themes and sub themes was developed. We used semi-structured interviews with the research team and other stakeholders from the project team, including collaborating commercial organizations. We also conducted observations in situ at project meetings, where collection and management of the data was discussed, and ran a focus group with the team to validate and discuss the main findings. We had full access to project documentation as well as the wider literature in the field, which was reviewed as part of the larger case study report for the Digital Curation Centre, upon which this paper is based.

Coding and Analysis

Transcripts were transcribed and inductively analysed as soon as possible after the interview so that emergent themes and concepts could be incorporated into subsequent interviews, and thus inform the direction of the research. The text of the transcripts was then cross-matched to the themes to provide systematic and transparent evidence of the research process. Emerging themes were reviewed and examined in relation to existing literature and theories, and in addition compared across datasets in the full study, the key points of which are summarised here.

Semi-Structured Interviews

Interviews were conducted after obtaining consent to provide an in depth exploration of the issues arising, as perceived by participants, regarding Telecare data. A topic guide was created and emergent themes from ongoing analysis were fed back into subsequent interviews to revise the topic guide, which explored issues such as the roles of researchers, the issues surrounding data collection, cleaning, analysis and storage and finally, the practical and policy dynamics of storage and curation, including funder protocols. Specific questions were targeted towards principle investigators and researchers, reflecting the issues they faced with tasks they undertook.

Participant Observation

One of the researchers attended the weekly Telecare team meetings for a year and a second researcher joined for a series of meeting during the study reported here. They observed team dynamics to gain an understanding of the issues faced by research teams within the context of carrying out the studies in question.

Focus Group

We conducted a focus group as a means of validating, refining and taking forward the issues arising from interviews and observation. Focus group participants were also asked to collaborate in mapping the interfaces they have with other relevant groups in the care management process to highlight less visible dependencies and affordances, and to identify gaps, overlaps, duplications and critical interfaces.

Participant Sample

We used a snowball sampling approach to achieve maximum variation and diversity of experiences. Our final sample included four primary investigators – three of whom were GPs involved in a range of telehealth pilots projects, three programme managers, three researchers, two telehealth centre call staff, two IT/data managers (one from the participating IT service provider and one from the NHS), and one researcher on data management from the Scottish Centre for telehealth. We also draw on the experience of the patients, nurses and GPs in an earlier qualitative study with this team (Ure et al., 2009a) as a secondary source of information about the factors impacting on data quality.

The Social Life of Tele-Monitoring Data

The first section outlines the issues of collection, processing, analysis and use of tele-monitoring data in clinical trials. This is followed by a discussion of the challenges and the opportunities for curating such data federation that arose from interviews with those in the field. Figure 2 provides some context for understanding the data lifecycle in a typical tele-monitoring study.

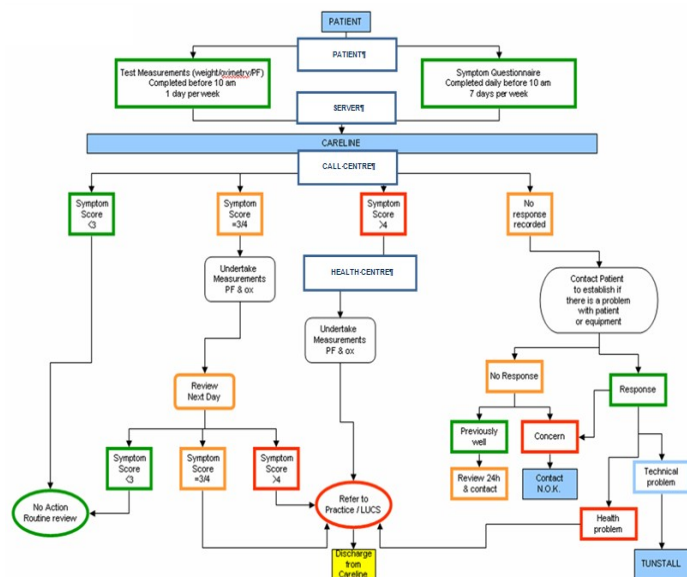


Figure 2. An early draft protocol for decision support with COPD monitoring data.¹

¹ “Oximetry/P” refers to measurement of respiratory rate. “LUCS” refers to the regional out-of-hours call out service.

The decision-support protocol in Figure 2 is an early draft version, illustrating in more detail the typical lifecycle of tele-monitoring data, involving multiple transformations of data in transit from (i) digital sensors at home, through to (ii) storage in a secure database and transformation by an algorithm prior to encryption for transmission to a call centre, then (iii) de-encryption for selective processing using a decision-support protocol to identify those patients requiring intervention, and (iv) transmission to the named contact at the health centre who will initiate either a call to provide more information, or a visit to the patient. One of the particular challenges identified for data curation in telehealth is to account for the very distributed (and often invisible) process of data shaping (Duguid & Brown, 2000) and data sorting (Bowker & Star, 2000) in transit across multiple constituencies from home to health centre. We look in the following sections at the issues which arise, and the implications for data curation in such contexts. These staging posts provide a focus for understanding the nature of data curation issues and strategies in tele-monitoring.

(i) Data Collection/Generation by Patients (Home)

Data from wireless peripherals depend on patients using them correctly to take and send data on vital signs and symptoms, and on the robustness of the technology and the telecommunication connections that transmit them. Any data curation process must take account of the quality and context of data collection in the lifecycle, and both interviews and observation demonstrate many sources of (often unanticipated) bias, omission or distortion in this very distributed data flow. The so-called “social life of data” (Duguid & Brown, 2000) is clearly in evidence here. The failing battery in a peripheral, the re-calibration of a scanner, the difficulty of using equipment, such as digital blood pressure cuffs, at home and the errors in data input were not evident to the receiver. They only became evident where multiple parties came together as part of the research process, or where the nature of anomalies were sufficiently extreme to raise questions about the validity of the data. One of the issues that becomes evident is the extent to which data quality is hard to monitor in a very distributed, and often automated process, at scale. Patients in an earlier qualitative study of tele-monitoring described at times how they were able to select which of their scores to transmit for example, to achieve (or avoid) a particular outcome, such as hospitalisation. This was not necessarily evident to the person at the end of the chain. Nurses with knowledge of the particular patient and their circumstances were often essential to interpret such data accurately. Similarly, technical factors impacting on the data were not evident to receivers of that data. If no data is received, does this mean the patient is unable to transmit their readings, or that the system is not transmitting properly? The role of these “unknown unknowns” in the quality of the data (and therefore of research or care) is substantiated in other recent white papers arising from attempts to share HealthGrid data between centres in UK, EU and US projects. (Breton et al., 2005; Ure et al., 2007).

From a data curation perspective, in very distributed contexts, users themselves have a potentially valuable role as curators in ensuring both the quality and the validity of the data, and in sanctioning the use to which it is put, including re-use for other purposes not originally anticipated. Recent research with young people on the use of electronic health records (EHR) highlights a wish to have the kind of flexible control of data use that they are familiar with in other contexts such as social networking sites, and a greater awareness of the risks associated with data sharing (Paterson & Grant,

2010). Companies such as MyDex² increasingly highlight the potential of user-based information management services in cutting the cost and risk of ensuring the currency and quality of data, better leveraging the value of that data for users and other stakeholders, and also enhancing data governance and data use/re-use.

(ii) Transmission and Transformation (Secure Server)

After transmission to a secure server (sometimes those of a company in the first instance) the raw data is converted into more tractable form for clinical decision-making. This may be based on commercially sensitive software packages used for decision support in other applications. In the case of COPD, the algorithm was the subject of commercial copyright and not in the public domain, making documentation of this process difficult. In the case of the hypertension study, data was converted into scores in the server of the German company whose mobile phone software equipment was being used for the clinical trial, and then transferred to a secure NHS server in the UK. In addition to the ethical, legal and commercial implications of transmitting digital readings across multiple jurisdictions, there are difficulties in documenting processes and transformations that are not open to inspection. In part, this reflects the experimental nature of these pilots. However, there is clearly an inter-organisational element to mapping the provenance and transformation of data.

Data curation in this context would require agreements with healthcare companies participating in health research to maintain records of algorithms used, even if these were not available for inspection at the time of use, as in the case study. Given the speed and the cost of technical innovation in this area, there is also the issue that data collected with one system may not be directly comparable to data collected using a subsequent system. The cost of maintaining access to systems that are no longer commercially viable is an, as yet, unresolved issue that impacts on the maintenance of mobile care systems already in use, as well as on the curation of data derived from them.

(iii) De-encryption and Decision Support (Call Centre)

Data from the secure server was encrypted for transmission to the call centre with a patient ID to allow linkage to patient records. The call centre was then required to (a) de-encrypt, (b) reassemble data, and (c) apply a protocol to identify those patients whose scores require action. This might be alerting a GP or issuing a reminder to the patient to take medication. This was done under time pressure, and the interfaces were confusing and difficult to use at speed, leaving open the possibility of human error according to call centre staff.

(iv) Interpretation and Validation (Health Centre)

The initial vision for tele-monitoring was based on the analysis of monitoring data using standard benchmarks for automated (or at least non-specialist) decision-support with regard to referrals. This vision of seamless data-sharing, like the HealthGrid vision, was tempered by the reality of implementing the system on the ground. The standard benchmark scores intended to guide prevention and early clinical intervention in COPD were found, in practice, to be of limited use alone. Individual variance across patient scores on breathlessness and blood oxygen levels reflected a range of context and person-specific factors such as weather, co-morbidity, preferred outcomes, and

² MyDex is a community interest company piloting this approach with Councils in two London Boroughs.

patients' previous activity. This made scores a very poor guide for clinical use in primary care, and, by the same token, for future re-use, unless complemented by the local knowledge of the practice nurse and the patients themselves. This more collaborative, patient-centred approach to diagnosis and treatment was complemented, rather than led by, the data.

Arguably, many aspects of data quality and confidentiality require designing-in local users as part of the curation of health data (Kaplan & Brennan, [2001](#); McGilchrist et al., [2007](#)). Patients are increasingly stakeholders in the management of their own care, and in some contexts, of their own records, as in Microsoft Health Vault³.

We have looked at the implication for curation of such data, given the many contextual issues in its generation, interpretation and use. What are researchers and clinical trial managers views on how this might be achieved? What, if anything, is currently done to annotate or store this kind of data in these fairly typical projects?

Barriers and Opportunities for Data Curation

According to Edwards, the American National Science Foundation has:

“exhorted their grantees to collect and preserve metadata – a prescription that has for the same number of years been routinely ignored or under-performed. The metadata conundrum represents a classic mismatch of incentives: while of clear value to the larger community, metadata offers little or nothing to those tasked with producing it, and may prove costly and time-consuming to boot.”
(Edwards et al., [2008](#))

We interviewed the team who collectively ran and managed these two studies, with a particular focus on what was currently done, what they felt could or should be done in future, and what the barrier and opportunities might be.

Lack of Incentives/Lack of Resources

While most interviewees felt data curation provided value, they reiterated that this would be additional work for a short term project, as most research projects are, without any benefit for those carrying out the work, requiring additional resources and training which were not currently included in the budget. The data collected was not currently annotated in detail, except insofar as required by the funders, the demands of the clinical trials infrastructure which many telehealth trials use (see Figure 3), and the documentation required by the IRAS⁴ infrastructure ([n.d](#)) to document ethical and legal consent in clinical trials. This additional work was widely regarded as unlikely to be done without incentives to do so, and without additional resource and training. This is echoed in the literature (Faundeen & Oleson, [2007](#)).

³ Microsoft Health Vault: <http://www.microsoft.com/en-us/healthvault/track-manage/chronic-condition.aspx>.

⁴ IRAS Integrated Research Application System: <https://www.myresearchproject.org.uk/SignIn.aspx>.

Extending Current Process Infrastructure

Project Investigators agreed that an extension of these familiar procedures would help incorporate data curation into existing practice.

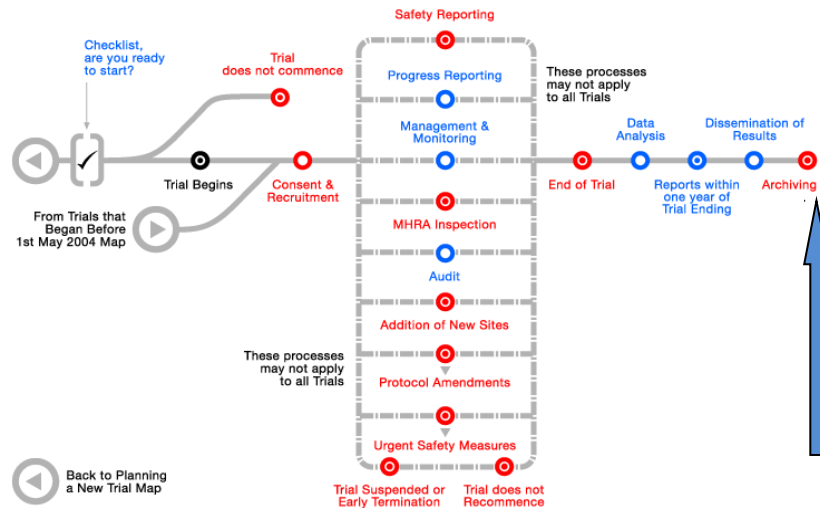


Figure 3. Building on the clinical trials infrastructure⁵.

Gallagher et al. (2009) also suggest (as did some of our interviewees) that “provisions for archiving should be built in at the proposal stage and consent should be designed around these provisions,” where such archiving was not intrinsically unethical.

A number felt the IRAS framework for managing ethical and other permissions could be usefully expanded to meet aspects of data management and governance at this stage. It is worth repeating in this regard that ethical and legal consent for use of trial data in the case study was sought only for the purposes of the clinical trial, and although some of this data can be re-used in anonymised form, consideration of the necessary permissions for re-use is something that could be flagged in IRAS as an issue for longer term curation prior to trial approval. Given that meaningful re-use may require access to patient records, and that consent is typically given for short term access for a specific purpose, by a specific group, this is a complicating factor.

Providing Incentives and Procedures from Funding Bodies

As Edwards et al. (2008) points out in an overview of infrastructure development in the US for the American National Science Foundation, incentives for curation are currently lacking, and this was clearly evident from the comments of our interviewees, who pointed to the difficulty of working to tight schedules, with limited budgets, and a range of unpredictable issues to address. In such a context, data curation implied a significant resources for which no budget was available, and for which expertise was seen as lacking. Compliance with requirements for providing metadata is typically reported as poor not only because of poor incentivisation, but also because systems are not always user-friendly, and healthcare staff need to be trained to be competent and comfortable using new systems (Whitten, 2006). Making these accessible and user-friendly is one strategy for mitigating this. Although recommendations from funders

⁵ Image taken from the Clinical Trials Toolkit: http://www.ct-toolkit.ac.uk/route_maps/map_landing.cfm?cit_id=248.

have been very general to date, there are now moves to develop more formal procedures, such as those embodied in the Data Discovery Gateway proposed in the Data Support Service project for re-use of data from projects funded by the Medical Research Council (MRC) to provide a user-friendly mechanism to define and publish the content of clinical research datasets, with online guidance on recognised standards and good practice in preserving and sharing data. It will also enable researchers to discover relevant MRC population-based datasets via a Web-based gateway to a catalogue. It is not a central repository of MRC-funded datasets however, and each research organisation will still remain responsible for the quality and security of its data and for decisions about collaboration and access. However, it will provide a vehicle for addressing some of the vexed questions raised by GPs and researchers about publication rights, where data from one study is re-used in other publications.

Incentives/Disincentives from Publishers

Journal publishers provide incentives to researchers in the requirements they make for submissions. Making data from studies available for secondary analysis and publication is one that could be instituted. For clinicians publishing results however, there was a concern that hard-earned data might simply then be open to re-use by others with little effort, acting as a disincentive. One interviewee suggested that this might be acceptable to her if the process required joint working by the authors of the different study to produce jointly authored papers. Gallagher et al. (2009) also suggest partial sharing of data and collaboration with original authors as a basis for better understanding of the data and to support future collaborations. This not unlike the approach adopted in some large commercial organizations such as BP, where expertise in increasingly shared through links to expert individuals, together with the data, as a more effective means of knowledge transfer where some information may be sensitive or may require a more in depth. Arguably, engagement between the research groups could facilitate a better understanding of the context of the data collection process, or the target group, given the evidence that knowledge of context, purpose and patient group is an important factor in data interpretation and re-use.

Data Access, Ownership, Management and Use

From the perspective of interviewees, the issue of re-use raised a number of grey areas, such as ownership and re-use of data, both in terms of who benefits from this and in terms of the risks associated with inappropriate re-use of data (access to patient records is often required to re-use the outcomes of clinical trials, for example). It is possible, however, to re-use the outcomes of the analysis of that data, or to have that data without identifiers, to minimize the likelihood that combination with other datasets will allow identification. There is a tension however, in that data linkage sufficient to allow for knowledge discovery in relation to disease also increases the possibility of knowledge discovery about patient identities. This tension is evident from exchanges in the British Medical Journal regarding the requirement for access to publicly funded research (Groves, 2009) where it is not clear that patient consent for use of their data in a clinical trial can then be re-used for some “unknown” future purpose (Greenhalgh, 2009).

From the perspective of older patients with chronic conditions interviewed in an earlier tele-monitoring study on precisely this issue, few had concerns about access to, and re-use of the monitoring data. On the contrary, their concerns were rather that changes in their condition might go unnoticed and prompt an avoidable crisis (Ure et al., [in press](#); Shipman et al., [2009](#)). However, there is evidence that younger patients, more familiar with sharing digital data using different tools, are more aware of risks, more concerned about them, and prefer more flexible control of this, along the lines evident in social networking sites (Paterson & Grant, [2010](#)).

The use of user-managed access is a growing response to these challenges in the context of health and government services for users (Maler, [2009](#)). It provides one solution to the unfolding complexities of data governance and confidentiality (Allan & Perkins, [2009](#); Bagdanovic et al., [2009](#)) in a complex, volatile and unpredictable digital landscape. Involvement of users in managing their own data is increasingly being seen to have cost and quality benefits, since users can curate their own data, and avoid the cost and duplication of operating on the basis of multiple databases with contact and content details in silos that do not correspond. User-managed access is now being adopted in commercial service developments with the general public, where cost, data quality and legal governance are at the greatest premium. Community interest companies such as MyDex⁶ provide accessible case studies of this (using the term personal information management systems) in the public service sector, but which have clear applications for health services.

The International Journal of Digital Curation
Issue 2, Volume 6 | 2011



The Challenge of Mobility

Mobile data was seen as exemplifying the complex security challenges of managing telehealth data in the scale, complexity and unpredictability of the problems that can arise, and the number of participants from temporary arrangements with commercial and academic organisations. In the case of the hypertension study, data crosses various jurisdictions where ownership and intellectual property, as well as access, are not clearly defined in law. In the hypertension study, the server for the commercial system being used for mobile monitoring is in Europe, and monitoring data is transmitted to a secure NHS server in the UK (in anonymised and encrypted form). In an ill-defined legal context, where agreements between stakeholders are short-lived, there is arguably a need for adaptive approaches that give users some representation in the ongoing process of managing access to their data.

This raises the question of how, where and when this can be integrated into the process of research, re-use and publication. Contractual obligations between organisations are harder where there are multiple changing commercial dependencies. These may not always be transparent, and legal and political frameworks do not respond rapidly to change. The potential of mobile systems to confer agency on individual users may provide some of the solutions here. Advances in software design in managing multiple distributed mobile users at scale suggests that alternatives are not only possible, but may have advantages in the potential to better leverage the local knowledge and agency of users themselves in providing information, or acting locally to improve it (Milner, [2009](#); Jami & Shaikh, [2008](#); Reddy et al., [2009](#); Latfi et al., [2007](#); De Toledo et al., [2006](#); Dabiri et al., [2003](#)).

Creating Community Infrastructure: Reconfiguring Digital Territories

The Data Curation Centre was viewed by two interviewees as a possible shared space to facilitate a representative process if incentives were provided by either funding bodies, or other academic or clinical stakeholders. The need for a forum, a facilitator and financial or other incentives is seen as an emerging priority in the strategic planning of patient groups and US healthcare policy strategy (US-ONC, [2008](#)). Patient groups and commercial companies, on the other hand, have already begun to demonstrate that with the advent of Microsoft HealthVault and similar infrastructure, that this may follow the pattern of distributed business systems, where the diversity of local knowledge and the agency of users is harnessed to mutual advantage in ways that top-down initiatives have failed to do.

Barriers and Opportunities to Engagement?

While the potential value was evident to users, an articulated vision of data integration and re-use for telehealth was not clearly articulated. From an individual perspective, one of the barriers to engagement with the issues – before incentives or procedures – was the opacity of the terminology within the digital curation community, and all of them commented on this. Opportunities for addressing common problems may be missed without the kind of brokerage that bridges cultural and semantic differences. The SCARP studies⁷ mentioned earlier highlighted the need for more direct bridges between professional communities. The need for shared language and shared spaces for exchange were evident in all the discussions. Interviewees felt that an organization such as the DCC could take a role in providing a shared space for

⁷ SCARP Project: <http://www.dcc.ac.uk/projects/scarp>.

← this kind of inter-organisational discussion and brokerage with stakeholders, as the UK eScience Centre did in the context of data integration in HealthGrids. →

From an inter-organisational organizational perspective, it seemed likely to interviewees that the clinical trials arm of this could extend existing procedures. Given the established national procedures for clinical trials, where data preservation and re-use is increasingly expected as part of the research, most interviewees felt this could be integrated and incentivised through the extension of existing infrastructure, such as the IRAS database. More internationally, this could also build on the work of EU-wide clinical trials initiatives, such as the ECRIN project⁸, and possibly the European Advanced Infrastructures in Translational Medicine or EATRIS project⁹. HealthGrid and Biobanking communities have something to contribute here in their experience of managing the representation of data across sites, scales and over time. Previous road mapping workshops in this context highlighted very typical problems and useful strategies for addressing them, including (a) collaborative metadata and ontology development consortia sharing tools, content and standards (UK eScience Centre, 2006), (b) road mapping opportunities to build on core measures of symptoms (Breton et al., 2005) and (c) use of early prototypes and pilots as a vehicle for generating engagement and exploring data quality. There is scope for re-use rather than re-invention here also if opportunities are provided for this to happen.

Emerging Strategies in Other Contexts

This paper drew on issues raised in the literature, such as the Data Curation Lifecycle Model (Higgins, 2007) and reports by HEFCE (2007), JISC (2007), OECD (2007), RIN (2007), EC (2007), Green et al., (2008), Martinez-Uribe, (2008), Beagrie (2007), Lavoie (2004), Lyon (2010, 2007), and the Interim Report of the Blue Ribbon Task Force (2008). It also considered research and development in other contexts where similar challenges arise in the eScience community (Edwards et al., 2008), the eHealth community (Breton, 2005; Ure et al., 2007) and the digital economy (Sawhney & Parikh, 2001; Tapscott et al., 2006).

Perhaps the biggest laboratory of strategy in this regard will be in the rapid development of national telemedicine networks in the developing world, where mobile telehealth networks are increasingly the norm, often using Open Source software. This is currently generating a whole ecosystem of models of moderated use and re-use that is arguably under-represented in the literature, given the novelty and speed of adoption, yet which represents mobile patient and community data management on a vast scale. The Harvard-based Sanamobile¹⁰ project providing Open Source software on a mobile phone for use in remote and rural areas in the developing world provides a good example of how this is being used.

⁸ ECRIN: <http://www.ecrin.org>.

⁹ EATRIS: <http://www.eatris.eu/Partners.aspx>.

¹⁰ Sanamobile: www.sanamobile.org.

Conclusion

Telehealth as a Laboratory for Strategy

As telehealth pilots now move into service development nationally, decisions will have to be made about the roles and rights of users and other stakeholders in this radical reconfiguration of the healthcare landscape. Much of the challenge for curation now derives from the need for organisations such as the DCC to provide opportunities and incentives for these communities to engage in the process of renegotiating roles, and rights of access in the context of use and re-use on which data curation will depend. Some of this can be incorporated into the existing infrastructure that supports applications for funding, for ethics, for clinical trials management, and as part of publication agreements in reputable journals. However, for many aspects of curation, particularly in relation to data quality and data re-use, the current approach is constrained by a lack of resources and incentives. This undermines the huge investment in generating tele-monitoring data as a unique new research data resource for machine learning to model trends and individualise treatment.

The DCC vision statement states that: “the scientific community has data characterised by structure, volatility and scale. Mobile tele-monitoring data offers a view of the kinds of data curation issues raised by highly dynamic and distributed systems in practice. These arguably require us to extend our notions of curation.”¹¹ . The current constraints on cost, combined with the scale and complexity of mobile tele-monitoring data, have provided a catalyst for re-thinking current curational frameworks, and the potential of patients as co-producers of quality (Kaplan & Brennan, 2001; Kaplan & Litewka, 2008; Allan & Perkins, 2009; Paterson & Grant, 2010), as patients increasingly have access to software to manage their own records.

References

- Allan, A., & Perkins, E. (2009). Adaptive access control emerges. Retrieved August 11, 2011, from <http://www.gartner.com/DisplayDocument?id=1124812>.
- Bagdanovic, D., Crawford, C., Coles-Kemp, L., (2009). The need for enhanced privacy and consent dialogues. *Information Security Technical Report 14*, 167-172. [doi:10.1016/j.istr.2009.10.011](https://doi.org/10.1016/j.istr.2009.10.011)
- Beagrie, N. (2007). *e-Infrastructure strategy for research: Final report from the OSI Preservation and Curation Working Group*. National e-Science Centre. Retrieved on December 10, 2007, from <http://www.nesc.ac.uk/documents/OSI/preservation.pdf>.
- Blue Ribbon Task Force on Sustainable Digital Preservation and Access (2008). *Sustaining the digital investment: Issues and challenges of economically sustainable digital preservation* (Interim report to the NSF). Retrieved August 11, 2011 from http://brtf.sdsc.edu/biblio/BRTF_Interim_Report.pdf

¹¹ “What is Digital Curation?” at: <http://www.dcc.ac.uk/about/what/>.

- Bowker, G.C. & Star, S.L. (2000). *Sorting things out: Classification and its consequences*. Cambridge, MA: MIT Press.
- Breton, V., Dean, K., & Solmonides, T. (2005). The HealthGrid white paper. In T. Solomonides, R. McClatchy, V. Breton, Y. Legre. & S. Norager (Eds.) *From Grid to HealthGrid*. Amsterdam: IOS Press.
- Consultative Committee for Space Data Systems (2002). *Reference model for an Open Archival Information System (OAIS)*. Retrieved August 2, 2011, from <http://public.ccsds.org/publications/archive/650x0b1.pdf>.
- Dabiri, F., Massey, T., Noshadi, H., Lin, C.K., Tan, R., Schmidt, J. & Sarrafzadeh, M. (2009). A telehealth architecture for networked embedded systems: A case study in in-vivo health monitoring. *IEEE Transactions on Information Technology in Biomedicine*, 13, 351-359. [doi:10.1109/TITB.2009.2013248](https://doi.org/10.1109/TITB.2009.2013248)
- De Toledo, P., Lalinde, W., Del Pozo, F., & Jimenez-Fernandez, (2006). Interoperability of a Mobile Health Care Solution with Electronic HealthCare Record Systems. In *Proceedings of the 28th IEEE EMBS Annual International Conference*. New York. Retrieved on August 2, 2011, from <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=04462980>.
- Denzin, N.K., & Lincoln, Y.S. (2000). *Handbook of qualitative research* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Duguid, P. & Brown, J.S. (2000). *The Social Life of Information*. Boston, MA: Harvard Business School Press.
- Edwards, P. et al., (2008). *Dynamics and tensions in eInfrastructure* (NSF Workshop Report). University of Michigan.
- European Commission (2007). Access, dissemination and preservation. Communication from the Commission to the European Parliament, the Council and the European Economic and Social Committee on Scientific Information in the Digital Age. Commission of the European Communities.
- Faundeen, J.L. & Oleson, L.R. (2007). Scientific Data Appraisals: The Value Driver for Preservation Efforts. In *Proceedings of PV 2007 International Conference*. Oberpfaffenhofen/Munich, Germany.
- Gallagher, M.D., Worth, A., & Sheikh, A. (2009, April 2). The challenges of data sharing for qualitative health research. *British Medical Journal*, 338.
- Green, R, Awre, C., Bayliss, S., Dolphin, I., Dow, M., Look, H. (2008). *RIDIR project: Final report to JISC*. Retrieved August 2, 2011, from <http://www.hull.ac.uk/ridir/Documents/ridir-final-report.pdf>.

- Greenhalgh T. (2009, April 14). Sharing medical research data: Whose rights and who's right? *British Medical Journal*, 338. [doi:10.1136/bmj.b1499](https://doi.org/10.1136/bmj.b1499)
- Groves, T. (2009, March 25). Managing UK research data for future use. *British Medical Journal*, 338. [doi:10.1136/bmj.b1252](https://doi.org/10.1136/bmj.b1252)
- Hanley, J., Ure, J., McKinstry, B., Pagliari, C. et al. (2009) Self-care at home: User and carer's experiences of tele-monitoring in the context of COPD. In *Proceedings of the NIHR SDO and HSRN Conference: Delivering Better Health Services*. Birmingham, UK.
- Higgins, S., (2007). Draft DCC curation lifecycle model. *International Journal of Digital Curation*, 2(2), 82-87. Retrieved August 2, 2011, from <http://www.ijdc.net/index.php/ijdc/article/view/46>
- HEFCE (2007). *Higher Education Funding Council for England strategic plan 2006-11*. Retrieved August 2, 2011, from http://www.hefce.ac.uk/pubs/hefce/2007/07_09/.
- IRAS Framework (n.d). Retrieved August 2, 2011, from <https://www.myresearchproject.org.uk/SignIn.aspx>.
- Irshad T., & Ure J. (2009). *Clinical data from home to health centre: The telehealth curation lifecycle* (SCARP Case Study No. 3). Edinburgh: Digital Curation Centre. Retrieved August 1, 2011, from http://www.dcc.ac.uk/webfm_send/106.
- Jami S. & Shaikh Z. (2008). A multi-agent-based architecture for data provenance in semantic grid. In S.A. Ao, A. Castillo, C. Douglas, D. Dagan Feng and J.A. Lee (Eds.) *Proceedings of the International Multi-Conference of Engineers and Computer Scientists (IMECS)*. Hong Kong: Newswood Publications.
- JISC (2007). *Digital Repositories Report* (JISC Circular). Retrieved August 2, 2011, from <http://www.jisc.ac.uk/media/documents/funding/2007/11/jisccircular507digitalrepositories.doc>.
- Jaegersberg, J. & Ure, J. (in press). Barriers to knowledge sharing and stakeholder alignment in solar energy clusters: Learning from other sectors and regions. *Journal of Strategic Information Systems*. [doi:10.1016/j.jsis.2011.03.002](https://doi.org/10.1016/j.jsis.2011.03.002)
- Kaplan, B., Brennan, P.F. (2001). Consumer informatics supporting patients as co-producers of quality. *Journal of the American Medical Informatics Association*, 8, 309-316. [doi:10.1136/jamia.2001.0080309](https://doi.org/10.1136/jamia.2001.0080309)

- Kaplan, B., & Litewka, S. (2008). Ethical challenges of telemedicine and telehealth. *Cambridge Quarterly of Healthcare Ethics*, 17, 401-416. doi:10.1017/S0963180108080535
- Latfi, F., Lefebvre, B., and Decheneaux, C. (2007). Ontology-based management of the telehealth smart home. In C. Golbreich, A. Kalyanpur, B. Parsia (Eds.), *Proceedings of the OWLED 2007 Workshop on OWL: Experiences and Directions*. Innsbruck, Austria. Retrieved August 11, 2011, from <http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-258/paper42.pdf>
- Lavoie, B., (2004). *The Open Archival Information System Reference Model: Introductory guide* (DPC Technology Watch Series Report). Retrieved January 3, 2008, from http://www.dpconline.org/docs/lavoie_OAIS.pdf.
- Lyon L., Rusbridge C., Neilson C., Whyte A. (2010). *DCC SCARP disciplinary approaches to sharing, curation, reuse and preservation* (JISC Final Report). Retrieved August 11, 2011, from <http://www.dcc.ac.uk/sites/default/files/documents/scarp/SCARP-FinalReport-Final-SENT.pdf>.
- Lyon, E., (2007). *Dealing with data: Roles, rights, responsibilities and relationships* (UKOLN Report). Retrieved January 3, 2008, from http://www.jisc.ac.uk/media/documents/programmes/digitalrepositories/dealing_with_data_report-final.pdf.
- Maler, E. (2009). The design of everyday identity. *Online Information Review*, 33, 443-457. doi:10.1108/14684520910969899
- Martinez-Urbe, L., (2008). *Scoping digital repository services for research data management: Project plan*. University of Oxford. Retrieved April 20, 2008 from <http://www.ict.ox.ac.uk/odit/projects/digitalrepository/docs/DigRepoProjectPlan.pdf>.
- Masis, V.G., Afsarmanesh, H., & Hertzberger, L.O. (2006). An agent-based federated information system for telecare environments. In *Proceedings of Information technology and Applications in Biomedicine (ITAB)*. Corfu, Greece.
- McGilchrist, M., Sullivan, F., & Kalra, D. (2007, December 13). Assuring the confidentiality of shared electronic health records. *British Medical Journal*, 335, 1223. doi:10.1136/bmj.39421.544063.BE
- McKinstry et al. (2009). Putting telecare to the test: The Lothian randomized controlled trials. In *Proceedings of Working Together Telecare and Telehealth Conference: The Scottish Centre for Telehealth Conference*. Stirling, Scotland.

- Milner, R. (2009). *The space and motion of communicating agents*. Cambridge University Press: Cambridge, UK.
- MyDex (2010). *The case for personal information empowerment: The rise of the personal datastore*. Retrieved August 2, 2011, from <http://mydex.org/wp-content/uploads/2010/09/The-Case-for-Personal-Information-Empowerment-The-rise-of-the-personal-data-store-A-Mydex-White-paper-September-2010-Final-web.pdf>.
- OECD (2007). *Principles and guidelines for access to research data from public funding*. Retrieved December 20, 2007 from <http://www.oecd.org/dataoecd/9/61/38500813.pdf>.
- Pagliari C., Detmer D. and Singleton P. (2007). *Electronic Health Records: Emergence and Implications for the UK* (Nuffield Trust Report). Retrieved August 12, 2011, from <http://www.nuffieldtrust.org.uk/publications/electronic-personal-health-records-emergence-and-implications-uk>.
- Paterson, L., & Grant, L. (2010). *Privacy and prejudice: Young people's views on the development and use of electronic patient records*. Royal Academy of Engineering. Retrieved August 2, 2011, from <http://www.raeng.org.uk/eprviews>.
- Reddy, M., Paul, S., Abraham, J., McNeese, M., DeFlitch, C., & Yen, J. (2009). Challenges to effective crisis management: Using information and communication technologies to coordinate emergency medical services and emergency department teams. *International Journal of Medical Informatics*, 78, 259-269. doi:10.1016/j.ijmedinf.2008.08.003
- Research Information Network (2007). Research funders' policies for the management of information outputs. Retrieved August 11th, 2011 from <http://www.rin.ac.uk/our-work/research-funding-policy-and-guidance/research-funders-policies-management-information-outpu>.
- Sawhney, M. & Parikh, D. (2001). *Where value lives in a networked world*. Harvard Business Review. Boston, MA: Harvard Business School Publishing.
- Seidel, E. (2009). Science and the socio-technical vision of cyberinfrastructure. In *Proceedings of the 2nd Communia Conference 2009*. Torino, Italy.
- Shipman, C., White, S., Gysels, M. & White, P. (2009). Access to care in advanced COPD: Factors that influence contact with general practice services. *Primary Care Respiratory Journal*, 18, 273-278. doi:10.4104/pcrj.2009.00013
- Strauss, A. & Corbin, J. (1990). *Basics of qualitative research techniques and procedures for developing grounded theory* (2nd ed.). Sage: London, UK.

- Tapscott, D. & Williams, A. (2006). *Wikinomics: How mass communication changes everything*. New York, NY: Penguin Group.
- Ure, J., et al. (2007). Data integration in eHealth: A domain/disease specific roadmap. In *Proceedings of HealthGrid 2007*. Amsterdam: IOS Press.
- Ure, J., Hanley, J., Pinnock, H., Kidd, G., McCall Smith, E., Tarling, A., Pagliari, C., Sheikh, A., MacNee, W. McKinstry, B. (in press). Piloting tele-monitoring in Chronic Obstructive Pulmonary Disease: A mixed methods exploration of issues in design and implementation. *Primary Care Respiratory Journal*.
[doi:10.4104/pcrj.2011.00065](https://doi.org/10.4104/pcrj.2011.00065)
- Ure, J., & Jaegersberg, G. (2005). *Invisible architecture: The benefits of aligning people, processes and technology: Sociotechnical case studies for system designers and managers*. Swindon, UK: British Computer Society.
- Ure, J., Kydd, G., Tarling, A., McKinstry, B., Hanley, J., Pagliari, C., Pinnock, H., Sheikh, A., McNee, W. (2009a). *Reconfiguring care: A qualitative overview of telemetry supported care for COPD patients in Lothian region* (Final Report for Intel and Scottish Centre for Telehealth). Edinburgh University Centre for Population Health Sciences.
- Ure, J; Procter, R., Lin, Y., Hartwood, Anderson, S., Lloyd, S., Wardlaw, J., Gonzalez-Velez, H. & Ho, K. (2009b). The development of data infrastructures for eHealth: A socio-technical perspective. *Journal of the Association for Information Systems*, 10(5). Retrieved August 2, 2011, from
<http://aisel.aisnet.org/jais/vol10/iss5/3>.
- US ONC (2008). US-ONC coordinated federal health information technology strategic plan 2008-12. Retrieved August 2, 2011, from
<http://www.hhs.gov/healthit/resources/reports.html>.
- UK eScience Centre (2006). *Proceedings of the UK eScience Centre Workshop on Data Quality/Usability: Issues in eHealth Practice Process and Policy*. Edinburgh, UK. Retrieved August 12, 2011, from
<http://www.nesc.ac.uk/esi/events/684/>.
- Whitten, P. (2006). Will we see data repositories for telehealth activity in the near future? *Journal of Telemedicine and Telecare*, 12(Suppl. 2), 65-71.
[doi:10.1258/135763306778393180](https://doi.org/10.1258/135763306778393180)