

Report on the NCeSS Agenda Setting Workshop on Confidentiality and Data Sharing

John Gekas, Udo Kruschwitz
Department of Computer Science
University of Essex

Hershbinder Mann, Dan O'Neill
Department of Health and Human Sciences
University of Essex

Simon Musgrave
KPMG

Department of Computer Science
University of Essex
Technical Report CSM-434
ISSN 1744-8050

September 2005

1 Introduction

Health Service providers, policymakers and researchers are increasingly interested in acquiring and joining data sources in order to get a more complete picture of service needs and social phenomena. This opens new challenges regarding patient confidentiality, data security (e.g. when transferred on a network) and secondary use of data.

The National Centre for e-Social Science and the University of Essex organised a workshop to explore issues arising from attempts to widen the usage of clinical and administrative data for research and planning, while protecting confidentiality and citizens' rights. The field of Learning Disabilities (LD) was chosen as it exemplifies many of the characteristics that could impede data sharing, notably that it is an inherently multidisciplinary and multi-agency field. The aim of the workshop was to engage a broad range of stakeholders who might be interested in identifying and overcoming obstacles to the service, policy and research usage of joined administrative and clinical data.

The workshop interspersed short presentations by domain experts with interactive discussions in the University iLab. The iLab comprises a room with a 360 degree white wall and PC workstations running 'brainstorming' software, and a breakout room with additional white board space, the use of which gave all participants an equal opportunity to contribute to developing an understanding of key issues.

Participants represented a cross-section of stakeholders, including:

- Learning disability service providers
- Policy makers
- Researchers
- Experts in web technologies
- Research Ethics Committees
- Advocacy services

The majority of the discussion concerned the potential benefits and problems that have to do with data sharing, particularly with regard to confidentiality. We also discussed possible solutions and future directions. This report summarises the discussions and findings of the workshop.

2 The Importance of Sharing Data

There is a fundamental assumption that there are significant benefits to be gained by sharing data. This section of the workshop asked participants for their views on what those benefits might be to them and to people with learning disabilities, who are the focus of the data. At the same time, participants were asked to consider whether data sharing is always important and desirable, and if not, why.

The majority of responses fell into two broad categories, *Provider Perspective* and *Patient / Client Perspective* with the former dominating (see Table 1 for a detailed breakdown). This was a little surprising, even though the workshop participants were mainly drawn from provider organisations, but perhaps underlines the necessity of consulting directly with patient or client groups in order to ensure that their viewpoint is adequately represented.

The overwhelming emphasis within the provider perspective was on the efficient and effective use of resources that follows from good planning and decision-making, whether at the level of the individual, the target population or for the translation of policy into strategy. This was seen as important for:-

- coordinating individual care
- understanding what is happening on the macro level
- reviewing whether the original policy objectives were 'correct' in the first instance

A second major concern was for integration and collaboration between providers. There were several references to gaps between services, for example comments about the desirability of '*joined up working*' and '*seamless*

CODES	Frequency of mention
Provider perspective	
Planning / decision making	41
Efficiency	37
Integration / collaboration	36
Quality	23
Added value	12
Research	14
Technical issues	10
Monitoring	7
Regulatory / legislative	2
Client perspective	
Power / influence	11
Openness / accessibility	2
Other	6
Social / community	2

Table 1: Importance of Data Sharing

services'. In addition, the potential benefits of sharing of experience and expertise, and avoidance of duplication of effort were all highlighted.

Integrated working was also seen as an important contributory factor to improving the quality of services provided, whether by “.. *support(ing) creation and implementation of 'best practice' based on real-life experience*”, or “*increasing capacity for 'joined up' policy & practice*”. Similarly, sharing data was seen as providing added value, principally through increased opportunities for learning within and across organisations. There were also several comments that expressed the sentiment that “*many heads are better than one*” - whether at the organisational level (“*Sound policy development likely to be dependent on different perspectives applied to common data*”) or the individual (“*no one is knowledgeable about everything*”). There was also the highly pragmatic view that “*we never collect quite what we need - but others might*”.

Several contributors clearly saw the potential benefits of data sharing for research purposes, for example having the “*ability to assess the effectiveness of a variety of clinical approaches*” or “*identifying issues of social equity/justice*”.

Interestingly, given that previous discussions with key informants have raised the possibility of barriers to data sharing arising from unclear legislation and guidance around the issue, there were few explicit references; there was only one reference to Caldicott, and another to the Freedom of Information Act. As stated earlier, the responses were very heavily biased towards the providers' viewpoint, and it might be that a client/consumer audience would have attached greater importance to legal controls and individual rights.

Finally, within the quiet voice of the clients, there were several comments concerning the potential for misuse of the information that might become available through data sharing. References were made to the power imbalance that could be created, which would favour those with resources, rights of access and understanding:-

“Because at the moment data is like an economic resource, control of it lies in the hands of a small number of people/organisations”

As a consequence it was seen as important that data should be made available to all “*to give more power to the user: can find data and services from different organizations*”. This is of particular relevance for the field of learning disabilities, as many of those who are affected would not have the capacity to identify potential misuse for themselves, nor to advocate for increased controls.

Finally, there was also the point that perhaps “*not all data needs to be shared*” and a reminder that data sharing could (or perhaps should?) be guided by the Caldicott principles, and that there should be a strict “*... control on sharing - (including) subject consent plus sharing on a need to know basis*”.

3 Data Sharing - Benefits and Concerns

Discussions were held in the “breakout room” adjoining the iLab in which each participant was encouraged to share their ideas about sharing data and to comment on those of others. It soon became clear that the ideas being articulated could be placed into one of two categories: on the one hand a core set of benefits was perceived and, on the other, a number of key concerns were expressed.

3.1 Benefits

There was agreement that data sharing was clearly beneficial with regard to informing and formulating policy. It was felt that the effective sharing of data would enhance ‘joined-up’ approaches and make management easier if different actors within the policy process could access the same information. In turn it was suggested that this would lead to efficiency gains by way of a more streamlined and integrated approach to decision-making as a whole. This was reflected by the desire of those in the workshop to see LD given prominence within the wider context of health care.

Participants were also keen to stress the benefits that would accrue to patients and a number cited improved care by way of using data to better distribute resources. One mentioned how services could be “*tailored to their [patients’] needs*” if the information necessary for this were freely available whilst others identified things that were more palpable: “*Ever been to a hospital that doesn’t have your notes?*” This rhetorical question highlighted something absurdly obvious yet fundamental to issues of data sharing.

Temporal aspects of planning were also heavily emphasised in the sense that data have a role in evaluating present and past policy before future policy is decided upon. Research and education feed into this process and participants agreed that more effective data sharing could only be constructive and positive.

Other contributions to the discussions linked data sharing to normative notions of democracy. One person in the workshop wrote of how data sharing can be “*fundamental for democratic projects*” and highlighted the “*equal distribution of power*” and “*balanced perspectives*”. More generally the group felt data sharing promoted equality and heightened accountability.

Finally, participants were very enthusiastic about the use of new technologies to address information concerns with regard to LD. In particular they perceived great potential in new technology to collect, store, search, and distribute data, and approved of the role technology could play in attaining the benefits set out above.

3.2 Concerns

The overriding concern of the workshop participants related to legal issues connected with consent and privacy. They were uncertain if individuals would be willing to consent to have their personal information linked between databases and whether it was legal for different sources to be linked even if consent was given.

These apprehensions spilled over into a wider issue of access control and the practicalities of granting access to legitimate users, at the same time securing data against misuse. The question of what authority is necessary to oversee a data sharing system also emerged. “*Who considers/decides what data are essential to pass on?*” was one query whilst others were founded on concerns about system security and permission to access data. Whether the confidentiality implications of holding information electronically might be different from those of holding physical paper records was also pondered by the group.

Data quality was another major area of worry with participants speaking of mistrusting data and a lack of consistency in the use of definitions, language and measures. Relating to this was the need for data to be as accurate as possible. This entails keeping records up to date and eliminating duplication which is not an easy task. “*Rubbish in, rubbish out*” was the warning of one contributor to highlight the necessity of effective quality controls at all stages from input to end-user.

Finally, the focus of the data was another issue that animated the workshop. The mismatch between what is collected and what is needed was a common complaint, as was the fact that data often focus on services rather than outcomes and benefits. In turn, this, along with some of the aforementioned issues, was linked to an apparent lack of incentives for data collectors to agree standards and actively participate in linking and sharing data. The multi-agency nature of the problem alone hinders the type of co-operation that is required.

4 Data Sharing Experiences

That there are many potential benefits that could follow from sharing data is unquestionable. Nevertheless, this fact on its own does not help us in assessing *how* data should best be shared. What follows is a summary of what participants thought has worked well or less well in their own experience in the context of sharing data. This session took place in the iLab.

4.1 What Worked Well

The points raised can roughly be summarised as falling under one or more of the following categories:

- Technical aspects
- Communication
- Organisational structures
- Data sources
- LD-specific issues

We will look at each of these categories in some more detail.

Technical aspects: Common standards in data files and explicit protocols for data sharing are important. However, equally important are security measures to prevent confidentiality breach as well as confidence in the security arrangements. A well-defined access control policy, e.g. a single point of access to an array of data sources (or services) is desirable which would allow access to data through something as simple as a web browser with a minimal amount of registration. *Semantic Web* technologies are seen as extremely valuable to social researchers to allow data to be browsed, searched, analysed and downloaded without having to send data on a physical medium, e.g. CD-ROM.

Communication: An important experience is that good communication between all parties allows for better information sharing. Personal relationships and access to data curators helps a lot. Part of a successful communication process is to explain what the data will be used for because an understanding of the potential benefits motivates sharing. This includes communicating with the actual data subjects who could clearly see what data was being gathered, why, and with whom it would be shared (this also leads to a straightforward and efficient consent process). Sharing data is seen to be easier if it is done for a specific purpose, for example direct care of the patient, or developing a commissioning strategy for services.

Organisational structures: Strong project leadership is key to success. Complex problems may be resolved at the individual level. Examples of successful projects were reported between NHS Trusts/clinicians; one salient feature of these projects was that collaborators shared a similar organisational culture. Greater difficulties arise when this degree of communality is absent.

Data sources: It was reported that freely available data sources can be useful. However, the type of data sources do not seem to be central to the success of certain approaches. This supports the idea that the nature of the major barriers to data sharing are more 'social' than 'technical'.

LD-specific issues: The discussion also brought up a number of aspects that are particularly relevant in the context of learning disability data. Complex LD needs or issues may *require* several types of information to be shared. In relation to vulnerable adults, procedures for sharing information are critical to effectively protect adults with LD (e.g. for supporting someone in court). Sharing data seems to work well if there is consent from the patients or "data subjects", or if the person with LD is in fact the one sharing his or her own information. Sharing works well when targeting specific individuals for improved care. Successful information sharing between professionals and other agencies has been evidenced in work concerning the protection of vulnerable adults from abuse.

Apart from the above issues we collected a number of practical experiences that have worked well. First of all, in demonstrator projects such as Electronic Health Record (EHR), if demonstrators took a pragmatic view of confidentiality issues (as opposed to overly legalistic), this generally meant that systems got off the ground and led to practical benefits to all concerned. Furthermore, where legislative barriers exist that effectively prohibit data sharing, informal, practical, information sharing arrangements between professionals (to get the job done) are seen to be in the best interest of the patient or client. Critical factors in these lower-level arrangements

are that the professionals concerned have a degree of mutual trust, and that they have the common goal of providing the best quality service for the patient or client.

In summary, sharing information works best if there are clear objectives about *why* information is to be shared and with *whom* because this helps decide *what* information needs to be shared and this will determine *how* this should happen.

4.2 What Worked Less Well

Obviously, there are a number of issues that make sharing information a less successful experience. We shall summarise the participants' concerns using the same categories as above.

Technical aspects: Things worked less well if there were inconsistencies in data coding or in basic definitions, incomplete data sources, difficulties with linkage of data (matching data from different sources), data format integration problems, and uncertainty about the validity/reliability of the data from the different sources. Another technical concern is data duplication, e.g. different departments duplicate the same data and do the same work twice. Changes in the data definition lead to further technical problems. To add another difficulty, different organisations use different terminology to describe similar concepts, and similar terminology to describe different concepts.

Communication: Agency leaders/directors who do not communicate ideas/requirements effectively can present problems, as can the lack of agreed procedures. Service, or policy-led approaches that do not work in partnership with the data providers are also more likely to founder. It is also essential to devise appropriate means of communicating with the client or patient. Concerns about how the combined data and information obtained will be used can then be addressed appropriately.

Organisational structures: It was reported in the previous section that projects within the same organisational culture can work well. On the other hand, teams with people coming from different cultures can work less well. Complex bureaucracy is another obstacle. Agreeing protocols/procedures between (culturally) dissimilar organisations, e.g. health and social services, is also seen as a problem. It is likely that this contributes to the development of low-level information sharing agreements, rather than robust organisational level procedures.

Data sources: It can be a problem to get hold of real data. This can be for a variety of reasons, including potential commercial value and confidentiality issues, but also personal investments and territoriality. It can even be a problem to get hold of recent data rather than data that is 2 years or more out of date.

LD-specific issues: An important question is whether people with learning disabilities are included and heard, and what language is used. Involving users has to be meaningful if trying to share data to improve services. Another question raised was whether the policies are being agreed on between the involved organisations alone, or whether the views of the patients/clients are taken into account as well. This will affect the success of the data sharing experience.

Apart from the above-mentioned problems, there are varying information sharing and confidentiality policies between different organisations. The reluctance to officially authorise data sharing between agencies or, to be more specific, the reluctance of those controlling data access, e.g. Caldicott Guardians, to sign off information sharing agreements are also considered to be problems. The same applies to defensive working practices that focus almost exclusively on risk (rather than potential benefit). Data restrictions can be frustrating. This is exacerbated if one has to enter into complex agreements and registration processes. Considering finer points of law/ethics are not usually effective in moving things forward - consulting the public and involving service users is generally far more productive in developing approaches that are effective and acceptable.

A major obstacle for sharing data can be people confusing the *why* with the *what* and the *how*. There can also be conflicts between the needs of users and providers. Finally, it is important to find a balance between individual rights and public good.

5 Specifying the System

In the last session all participants worked in small groups to come up with solutions that would join up different data sources leading to practical applications. In other words, if we are able to combine different data sources following strict confidentiality guidelines, what would we like to use them for?

A number of approaches were presented, among them work in progress such as the CLEF project and ongoing

work at Essex County Council to build an integrated system with access to a variety of data sources (like social and housing information) that can be used by staff who work for the council.

It was pointed out that any such system should only use relevant and up-to-date data and should clearly show where the data comes from.

Other interesting suggestions made were:

- Parents of people with learning disabilities consistently raise inaccessibility of up to date information about services as an issue. How could this situation be tackled?
- Users and carers could access the system with appropriate level of access control.

It was also noted that the investment for IT for learning disabilities and mental health was small in comparison to social and direct costs.

6 Conclusions and Future Directions

One interesting observation is that the discussion focussed less on confidentiality as such, but much more on the basic problem of sharing information in the first place, possibly indicating where the greatest concerns lie. An alternative interpretation is that the participants largely represented the viewpoints of the researchers, planners and providers of services, and as such were mainly interested in the practicalities of data sharing. Users of those services might have a different viewpoint, and this is something that we shall revisit in future.

We also conclude that a number of research projects are underway to:

- implement a common architecture that can be used within the NHS to improve patient care and services (e.g. the *National Programme for IT (NPfIT)* framework¹). However, data flow into the 'National Spine' will consist only of a small core of clinical data, and will not include data currently held in the many health and social care systems throughout the country. Consequently, in the short to medium term, the National Spine might be of limited value as a data source for research and planning purposes.
- make clinical data available to assist in the treatment of patients and to derive scientific data from a variety of data sources (e.g. the MRC-funded *Clinical e-Science Framework (CLEF)* project [3])
- use *Semantic Web* technology to produce joined up data (e.g. the EU-funded *Data, Information, and Process Integration with Semantic Web Services (DIP)* project²).

However, what is lacking is a framework that has the carer or the patient as a potential user of the services in mind. This is becoming increasingly important given the introduction in recent years of legislation designed to promote the inclusion of disabled people and other vulnerable groups, and to support the rights of the individual in society (e.g., the Disability Discrimination Act, Data Protection Act & Freedom of Information Act). All the confidentiality and data sharing issues apply equally in the context of learning disability services, but the situation is complicated further by the service users' reduced capacity to engage with the development of data sharing systems.

This sentiment is strongly supported by evidence reported elsewhere, e.g. in an Essex County Council commissioned study [2] for which disabled people were interviewed and which found for example:

“Lack of accessible information about Social Services, and the potential support and advice that may be on offer for disabled people, emerged spontaneously as a key issue of concern for many of the people we interviewed.”

There is further support for such an approach when we look at the exemplar domain of learning disabilities in particular. It has been reported that an estimated 60% of adults with learning disabilities live with their families [1]. These carers “*face many problems and challenges. They need more and better information*”. One of the specific challenges identified by the government paper [1] is to ensure that carers “*obtain relevant information about services*”.

¹<http://www.connectingforhealth.nhs.uk/>

²<http://dip.semanticweb.org/>

For the LD domain we conclude that, no matter what application we are looking at, the data sharing and confidentiality concerns are similar. However, we would like to see more efforts directed at developing services that join up data for the immediate benefit of patients/clients and carers.

References

- [1] DEPARTMENT OF HEALTH. Valuing People: A New Strategy for Learning Disability for the 21st Century. Government White Paper, 2001.
- [2] JOHNS, T., NICHOLAS, N., JOHNSTON, J., COOPER, G., CARTER, Z., AND MILLER, P. The equal lives evaluation report summary. Essex County Council Publication, 2004.
- [3] RECTOR, A., TAWHEEL, A., ROGERS, J., INGRAM, D., KALRA, D., GAIZAUSKAS, R., HEPPLER, M., MILAN, J., POWERS, R., SCOTT, D., AND SINGLETON, P. Joining up Health and BioInformatics: e-Science meets e-Health. In *Proceedings of the UK e-Science All Hands Meeting* (Nottingham, 2004).