GRADE

Scoping a Geospatial Repository for Academic Deposit and Extraction

JISC DEVELOPMENT PROGRAMMES

Project

Project Acronym	GRADE	**Project ID**	
Project Title	Scoping a Geospatial Repository for Academic Deposit and Extraction		
Start Date	1 June 2005	**End Date**	30 April 2007
Lead Institution	EDINA		
Project Director	David Medyckyj-Scott, EDINA		
Project Manager & contact details	Anne Robertson, EDINA University of Edinburgh Causewayside House 160 Causewayside Edinburgh EH9 1PR Tel 0131 651 3874 email a.m.robertson@ed.ac.uk		
Partner Institutions	University of Edinburgh, University of Southampton		
Project Web URL	www.edina.ac.uk/projects/grade		
Programme Name (and number)	Digital Repositories		
Programme Manager	Neil Jacobs		

Document

Document Title	Geospatial Data-sharing in UK Higher Education: informal repositories and users' perspectives		
Reporting Period	Nov 2006 – April 2007		
Author(s) & project role	Robin S. Smith, Informatics Collaboratory of the Social Sciences, University of Sheffield, 219 Portobello, Sheffield, South Yorkshire, S1 4DP. Project Consultant		
Date	April 2007	**Filename**	GRADE_reportRSSv2.pdf
URL	www.edina.ac.uk/projects/grade		
Access *	☐ Project and JISC internal	x General dissemination	

Document History

Version	Date	Comments
1.0	12/04/2007	Simplified SWOT analysis and Appendix 1 added
0.1	23/03/2007	First draft of workshop report

Smith, R.S. 2007 Geospatial Data-sharing in UK Higher Education: informal repositories and users' perspectives. Sheffield, UK: University of Sheffield, 64pp. (GRADE Project Deliverable WP2)

Contents

Abstract
This report outlines research into the sharing of geospatial datasets by researchers based in UK universities as part of the GRADE project, a scoping exercise for the creation of a geospatial data repository for UK Higher Education. The report contains an informal assessment of technologies used to typically share geographical information and contrasted these through an experiment with novel, informal peer-to-peer data-sharing technologies and the GRADE project's demonstrator repository. The study adopts a qualitative research approach to help explicate the issues that representatives from the GI community experienced in the context of the experiment and concerns and opportunities presented as a result. The analysis includes a discussion of materials presented at a one day workshop that brought participants together and a SWOT analysis of both the informal sharing methods and the GRADE demonstrator repository.

A list of ten recommendations is given towards the end of the report, highlighting a need to consider the wider context of a research data repository in terms of educating its user base in data policy and licensing considerations relating to GI; the need to continue and develop such repositories in relation to wider research and (national) geospatial data infrastructures; to adopt better practice with regard to metadata handling and creation for both the resource and the community using it; and to recognise the opportunities such a resource presents for both the development of technical tools to support research and a environment to support qualitative research into such activities, as part of a wider information society.

Acknowledgements
The author wishes to thank the participants of the information-sharing workshop where the majority of the material used in this report was derived. It should be noted that, although attempts have been made to readily reflect the opinions of the participants, any omissions or errors in the report are the author's alone.

1 Introduction

The Scoping a Geospatial Repository for Academic Deposit and Extraction (GRADE) Project[1] is one of a cluster of projects in the Digital Repositories Programme funded by the Joint Information Services Committee (JISC) of the Higher Education Funding Council for England (HEFCE) investigating the interactions between data and institutional (publications) repositories, support for scientific lifecycle, storage and access requirements.

JISC is developing a programme of work relating to digital repositories. Its aim is to bring together people and practices from across various domains (research, learning, information services, institutional policy, management and administration, records management, and so on) to ensure the maximum degree of coordination in the development of digital repositories, in terms of their technical and social (including business) aspects.

Increasing user-expectations and personal information management capabilities has meant that there has been a rapid growth in the extent to which 'informal' repositories (such as peer-to-peer technologies, wikis, blogs, e-portfolios) have become implicated in routine workflows. There is a requirement to understand how these repositories function, the role which they play in the wider information landscape and the extent to which such repositories may (or may not) be honouring digital rights. These trends also represent an opportunity to evaluate received wisdom with regard to formal repositories and to experiment with emergent technologies prior to *de facto* entrenchment of informal practices within the community.

This report contributes to work-package 2 of the GRADE project by helping to evaluate the extent to which these informal repositories are currently used to share geospatial data and further the goals of geospatial asset repurposing. The overall aim of the work-package is to ensure that geospatial data is exchanged within a context that permits end-user activities to (legitimately) exploit rights-asserted assets. The main approach of the reported research was to provide a 'live' sharing experience for the researchers so that they could experiment with peer-to-peer technologies designed to share datasets that, typically, e-mail cannot currently cope with. In addition, participants were encouraged to upload data, and consider this form of sharing alongside the GRADE demonstrator repository[2] and their current ways of sharing. It was felt that providing such a common experience of data-sharing would help participants to uncover their attitudes to sharing together in a focus group/ workshop environment involving several tasks.

The workshop involved presentations by all participants around their previous ways of sharing data, their experience from the experiments using informal

[1] http://www.edina.ac.uk/projects/grade
[2] http://gradedemo.edina.ac.uk/dspace/index.jsp

repository technologies and a brief SWOT analysis[3] of both these 'new methods' and the GRADE demonstrator repository. They were also asked to carry out group-based tasks in the workshop to respond to some of the opinions expressed in a questionnaire survey that was carried out with respondents from the UK university-based Geographical Information Systems (GIS) community[4], in terms of barriers to sharing and future opportunities. Importantly, the participants were not only asked to view these activities from a purely technical perspective but also the social, cultural, policy and organisational issues present in data-sharing in the GI Sciences (GISc).

To maintain their perspective of the issues and allow their voices to be expressed, this research and report adopts a qualitative methodology and approach. This includes a degree of ethnography, as the report's author was both a participant in the experiments using the new methods and the facilitator at the workshop. As such, it should be noted that the author's comments are contained within the quotations in this report (all of which are anonymised) but that this material has also been circulated to participants for further comment to attempt to remove any degree of misinterpretation or over-representation by the author. It is hoped that by doing so, an open process of sharing the development of these ideas has led to honest opinions about what opportunities and barriers occur between both informal and formal means of sharing geospatial data.

As such, the research is not designed to reflect a statistical sample of the UK academic community who use GIS but it is useful to note some general characteristics of those who took part. The thirteen participants involved were recruited from two main perspectives, those that had already had some involvement in GRADE (and were invited by the contractor) and some colleagues and friends in the GIS community from around the country. One dimension of interest was the varying locations of users and what impact this may have on sharing. As such, some preliminary activity took place at the University of Sheffield, with additional participants recruited from colleagues in Leeds, York and Manchester before turning to those from further away, including some participation from staff from EDINA (although final levels of participation varied in several cases). The gender mix in the overall group was not balanced but perhaps reflects the level of uptake of GIS in the community (4 female, 9 male). Eight of the participants were English, two Scots, one non-British European and two non-Europeans. Generally speaking, the group included two senior members of staff, four with lecturing positions, four who were researchers and three participants that had research support or technical duties too. Perhaps most importantly, several of the participants knew each other through previous meetings and other events, which was also felt to help foster the honesty and openness expressed in their presentations and comments in the workshop.

---

[3] i.e. Strengths Weaknesses Opportunities and Threats

[4] Part of which took place at the GISRUK conference in Nottingham 4th - 7th April 2006 see: http://edina.ac.uk/projects/grade/revised_GISRUK_Questionnaire2.doc

The remainder of the report is split into three main sections. Firstly, the following section outlines the activities that the participants were asked to do and the three different technologies used in the 'experiments' to uncover their opinions. The second section deals with the evidence provided at the meeting through the aggregation of the participants' presentations, comments and other workshop activities. The third section draws some of this discussion together across the workshop topics around the general issues involved in sharing and the particular issues coming from the SWOT analysis.

Notably, the report lacks theoretical detail, with the exception of considerations towards grounded theory (see Bryman, 2004) and the social shaping of technology (Woolgar, 2003). One element that will emerge through the discussion is the impact that the study has had on the participants in terms of 'action research', as participants experienced a great deal of learning-by-doing, helping to inform the opinions expressed in the report and their future approaches to this topic. As such, in as much as the GRADE project is a scoping exercise for the potential to establish a repository for the UK academic community, this report presents a 'scoping' of the issues present in the potential user-community of that resource in relation to informal repositories and an indication to which theoretical approaches may be appropriate for further investigation.

2 Technologies and Tasks

In order to expose the participants' views about sharing GI, it was decided to conduct a brief experiment using fairly new but readily available Internet-based technologies as 'informal repositories', mainly utilising peer-to-peer features. Following some tests of several products (with varied suitability) two were chosen for their contrasting functions. A brief discussion of them is given below, with Appendix 1 illustrating some other examples.

2.1 A Plug-in Approach

AllPeers[5] is a free (beta version) plug-in for Mozilla's Firefox web browser. The tool allows you to invite 'friends' to various groups (e.g. "friends", "family" and "GRADE"), so that you can share data using a peer-to-peer approach. Icons for each 'friend' allow the user to see who is online and signed-up users can chat to each other through an instant-messenger (IM) function. Any file on the local machine can be shared with friends and AllPeers is designed to help people send files that would normally be too large for e-mail (due to problems with server capacity or firewall restrictions etc.). AllPeers was chosen for this experiment because it allows you to select who you want to share data with and that Unix users could also equally participate in the

[5] http://www.AllPeers.com

experiments. Friends could also be notified through e-mail and IM if new files could be accessed. The interface of AllPeers can be seen below (see Fig. 1: AllPeers- Navigator and Inbox).

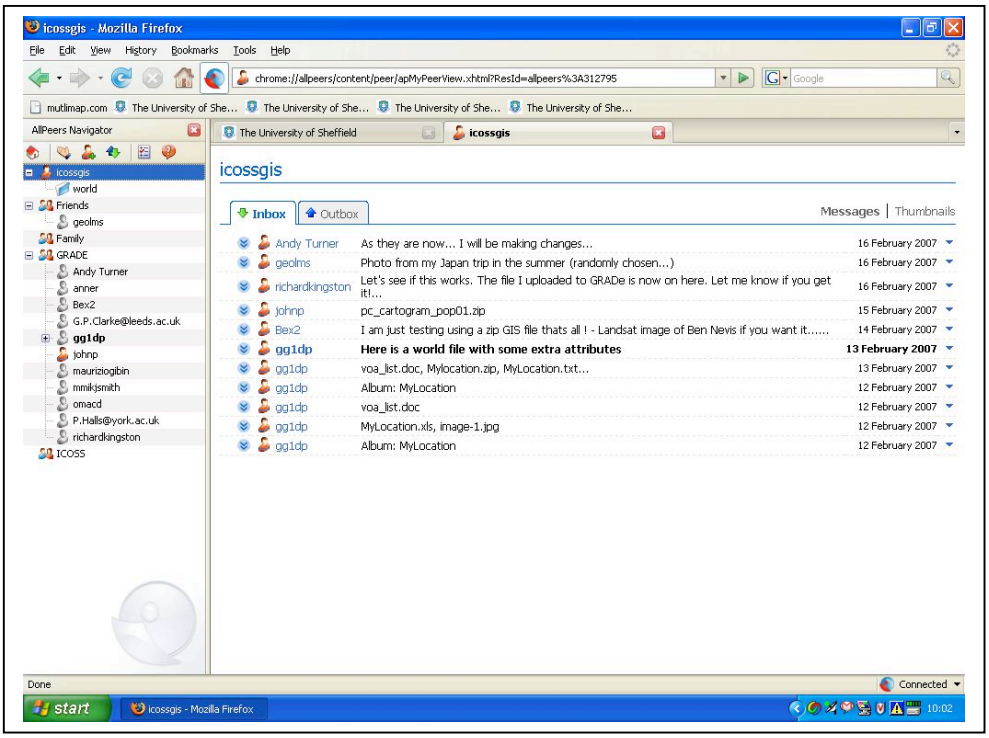


Fig. 1: AllPeers- Navigator and Inbox

2.2 A Client Approach

Exaroom[6] is a (beta version) portal-based sharing environment which involves connecting to a webpage to configure your settings and also invite 'friends'. It also involves installing a 'client' that installs "Ground Control"- their file-sharing software. Exaroom also allows you to see who is online but you can also passively browse a 'public' folder that other users have setup on their machines. This informal repository technology also promotes the ideas of users having remote access to files on another of their own machines, by pulling files from, for example, a home-based machine at work. Again, friends could also be notified if new shared files could be accessed through mailing services. The portal already has one 'friend' present, "Major Tom", the avatar of the service, something that would raise particular issues with the participants (discussed below). Exaroom is different from AllPeers as the portal offers users the facility to add more information/personal details about themselves and a couple of the participants decided to do this. A user's 'web-presence' is illustrated below, along with the file-sharing interface of Ground Control and Share Manager (see Fig. 2: Exaroom Interfaces- Webspace and File-sharing (inset)).

[6] http://www.exaroom.com/beta n.b. this is the current homepage link

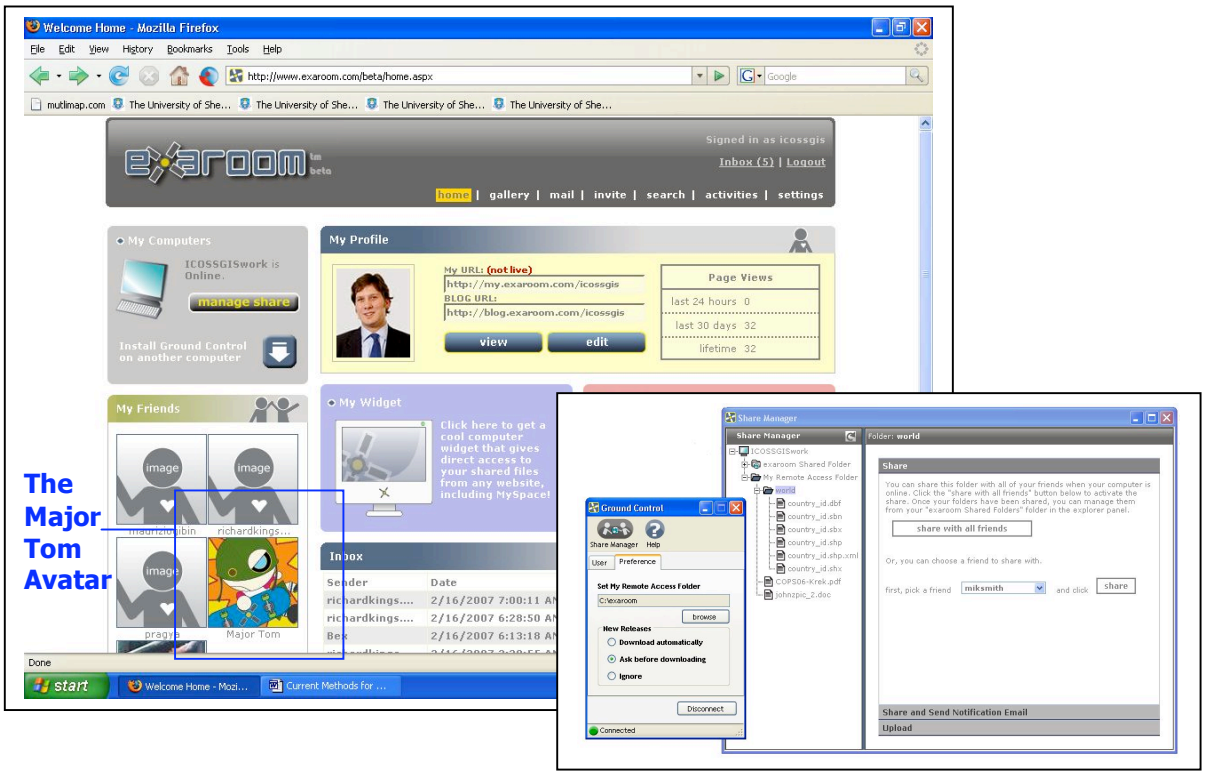


Fig. 2: Exaroom Interfaces- Webspace and File-sharing (inset)
(Major Tom Avatar highlighted)

The two technologies offered a couple of contrasting options:

- If the user did not want to (or could not) install Firefox (or the AllPeers plug-in) then Exaroom was available as a separate package
- AllPeers only allows users to add friends for direct sharing, whereas once connected to one friend, an Exaroom user could browse the friends-list of that contact, allowing them to explore a wider network (and potentially their shared files).
- Exaroom required people to be online in order for participants to passively pull and download data. AllPeers, for the most part, was employed as something that could be seen as more of a 'push' system.

2.3 Experimenting with the Informal Methods

It was intended that two experiments would take place with these technologies. The first was intended as a means to allow the participants to install the technology and share some data that could be readily added to, as a collaborative activity based on readily available GI. In this case, the experiment involved point data that represented the location of the participants' offices and 'attribute data' involving their contact details[7]. It was suggested that once data had been added, files would be passed in a pre-specified order creating a chain (in a 'pass-the-parcel' mode). The second experiment involved looking at the capacity of the systems to handle GI,

[7] A simple GIS operation that all participants would be familiar with and quick to do.

which frequently involves large datasets. A polygon dataset of worldwide coverage of countries (distributed by ESRI with their software) was set up with a simple numeric unique identification number for each record. This 'unique id' would then allow participants to add any attribute data as variables (either as text strings or numbers), although at least one participant added genuine data (e.g. national currency) during the exercise. It should be emphasised that, for both technologies, the aim was for the participants to have a common experience in sharing GI and to uncover a variety of social and technical issues that they were likely to experience as a result. To some extent it also illustrated the extent to which it was possible for these systems to share GI, which often comes as a collection of files rather than just the one that most media require, from spreadsheets to digital images.

2.4 The GRADE demonstrator repository

A third part of the experiment related to the GRADE demonstrator repository (based on Dspace Software[8]). In comparison to the informal repositories, this facility was seen as a common 'space' for registered users to browse and upload data, providing further details through some metadata fields and conditions/permissions for use. The interface of the repository demonstrator can be seen below (see Fig. 3: Searching and Submission for the GRADE Repository Demonstrator Homepage).

Fig. 3: Searching and Submission for the GRADE Repository Demonstrator Homepage

[8] see http://www.dspace.org

The GRADE demonstrator repository provided different information-sharing experiences when contrasted with Exaroom and AllPeers around four themes. The first difference is the asynchronous nature of the sharing experience, where participants did not need to interact to share data. Secondly, the repository was being exposed to a larger group of users, outside the confines of the informal repository experiments. Thirdly, it too provided a passive browsing experience but, unlike Exaroom, it was seen as less technical (as no additional installation was needed). Fourthly, the repository lacked the notification and communication tools present in the other methods, resulting in some participants choosing to notify the group that they had uploaded some data (by e-mail). This contact helped to establish increased interaction between participants before the face-to-face workshop took place.

In addition, by encouraging the group to use the GRADE demonstrator repository to share their own GI, a contrasting position exists with the two other experiments. For example, it was hoped it would help to expose issues relating to: data licensing, Intellectual Property/Data Rights, copyright and other data policy concerns, as this would involve sharing their own data. This data would, therefore, have required more effort to create and greater than the comparatively neutral value of the 'dummy' data used with the informal repository technologies. As such, these activities were not intended as a direct comparison of varying technologies for sharing geospatial data but, rather, the catalyst to expose the views of a small but engaged and active group of geographical information scientists.

3 Results (Part 1) Exploring Information-sharing

3.1 'Problems' in the experiment

After several days of preparation, the facilitator contacting participants and providing written guidance for all parts of the experiment, the activities went 'live' on the morning of 13/02/2007, with an intention that participants would find time to share data over the next two days. Within the first few hours, it became clear that the process had 'stalled' almost immediately, due to a mix of technical and, broadly speaking, social problems. In response, the facilitator suggested to the group that the experiment should be extended by an extra day but also specified a particular time to participate (as suggested by one of the participants during a one-to-one conversation). Unfortunately, the chain still kept breaking-down, to the point where it was suggested that (in the remaining time available) participants should try to share some data with another member of the group through the informal repositories and ensure that they uploaded data to the GRADE repository. Although, this may seem like the experiment 'failed', this is only the case in terms of the practical activity of 'chaining' the sharing of data, as all participants appeared to be

able to share data and contribute to informed and informative input at the workshop.

3.1 Current Methods for Sharing GI

The current ways in which the participants share GI was of interest to this part of the study in two ways. Firstly, it would show which methods the participants would normally consider to aid sharing and help them to reflect on such 'everyday' practices in relation to the experiments. At the start of the workshop the participants were asked to report their normal frequency ("frequently", "intermittently" and "infrequently") of sharing activities in terms of:

- who they would share with
- the location of that person or group
- the methods they would use

3.1.1 An overall pattern to sharing

The table below gives a generalised view of the different dynamics involved with sharing GI. The frequency of responses is indicated by the order of the responses in each cell but it should be noted that such responses are, in part, also influenced by the types of research the participants undertake. For example, if a participant frequently worked with external organisations then this method would be more prevalent (see Table 1: Frequency of Sharing: Who, Where and Which Methods). As such, statistical analysis of this position has limited value but further examination of the frequency of methods is considered below from a more qualitative perspective.

	Frequently	**Intermittently**	**Infrequently**
Who	Colleague/Team member Student/Trainee, Those with Internet access, Subscriber to specialist service	People elsewhere in HE Colleagues (locally) Students Public sector organisation	lecturers/colleague, students/trainees Outside HE
Where	Same room/lab, Same department, Same campus/university, The UK, "Global"	Across departments/same university overseas/ "beyond" lab/room Same locale	other (mainly UK) universities Same room/lab/office Across departments/ same university organisations in other locales, organisations in same locale
Methods	USB Flash Drive/ memory stick e-mail, CD-ROM, network drive/fileserver webpage/html FTP, Unix links, WebCT	e-mail, webpage, FTP, file server, CD, DVD (in post), USB/flash drive, Unix links SCP VKP	e-mail CD-ROM USB intranet shared directories/file server HTTP FTP

Table 1: Frequency of Sharing: Who, Where and Which Methods

3.1.2 Frequent sharing

The most frequent forms of sharing typically involve colleagues at the same institution and frequently the same room. USB storage devices are commonly used to share this data, closely followed by e-mail. Use of these two mechanisms was related to the volume of data involved or concerns about data being corrupted through e-mail delivery/receipt. Teaching, group or project work led to sharing taking place through online resources such as network drives and dedicated teaching environments (such as Web Course Tools- WebCT[9]). In contrast, however, one participant was eager to make their data more readily available to anyone who had access to the Internet through various web-based resources. Examples of participants' comments about frequent sharing situations can be seen from these statements (in terms of who, **where** and *the methods they used*):

"colleague- **same lab**- *USB stick, CDROM, personal webspace*"

"colleague, **UK**, *FTP*"

"Students; **Lab**; *Network Drive, UNIX Links*"

3.1.3 Intermittent sharing

Intermittent forms of sharing GI involved those who teach infrequently or who have project specific tasks, often with those at disparate locations in other universities or with, in a couple of instances, public sector colleagues. Some of this sharing was noted to be dependent on both the type of data involved and who they were sharing that data with. It is possible that the sharers' physical locations led to slightly different methods being employed. This activity also highlights the notion that sharing can involve more than just a dissemination process, as often found in the more local/frequent situation. For example, e-mail's popularity may be attributed not only to the sender's familiarity with (or ability to use) this technology but also the (perceived) familiarity/ability of the recipient to obtain the sent data. This activity also highlighted the adoption of specific technologies to aid sharing, including:

- the use of 'Secure Copy' (SCP) to share data from a local machine to remote hosts, and
- a project collaboration tool, Virtual Knowledge Park (VKP), to provide a dedicated shared resource between project partners.

Examples of participants' comments about intermittent sharing can (again) be seen in terms of who, **where** and *methods*:

"With colleagues; **overseas (Australia, USA)**; *email, FTP, DVD in post*"

"PhD students; **same room**; *flash drive/e-mail*"

[9] see http://www.webct.com

"People **elsewhere in HE**, *by email or FTP*"

3.1.4 Infrequent sharing

Participants were less able to identify infrequent times when they shared data. No new methods were uncovered in this group and the only main difference with intermittent forms of sharing seemed to be when participants chose to share with colleagues in other (mainly UK) universities. Again, the purpose of sharing (such as project work) seemed to be a driver to this activity, where sharing with partners outside the UK was less common than in the case of intermittent activities. Examples of infrequent use can be seen from these statements, again in terms of who, **where** and *methods*:

"Colleague, **same room**, usb"

"People **outside HE**, by email or FTP"

"Trainees, **different Universities**, e-mail and cd"

This topic also raised issues about instances when someone moves university. One participant outlined the case where a shared network drive with various large datasets was used with students but that these were too large to be burned to CD-ROM for the move to a new university. Subsequently, the datasets have been lost. This example may also lead to questions about ownership of a dataset between the individual and the institution and where such data should be stored for long-term curation[10].

It can be seen that participants readily identify different types of 'actor' to share with, that these may vary depending on the project context, that location plays some role in a number of these instances (particularly in terms of how people negotiate sharing) and that participants have identified several different methods, some of which were specific to these other relationships. Of course, this also indicates that the group involved in the project does share geospatial data on a regular basis, that some have quite sophisticated technology to facilitate this process but that e-mail, USB, 'burned media' (e.g. DVD-R) and shared network storage cover the majority of activities. Notably, nobody in the experiment mentioned being a user of the informal repository methods adopted for the experiments.

3.2 New Informal Methods for Sharing GI

During the workshop, participants were asked to outline what they felt the role of AllPeers/Exaroom could be compared to their 'normal' methods of sharing GI. This varied from those who felt that such informal repositories were too different to be of any use to potential roles and benefits that they appeared to offer.

[10] Recent activities at the Digital Curation Centre (http://www.dcc.ac.uk) relate to these concerns for the long-term maintenance and accessibility of GI.

3.2.1 Problems with the new informal methods

From a negative perspective, the applications were seen as ill-fitting:

> "AllPeers/Exaroom… are you kidding me?... Ground control application (again…r u ['are you'] kidding me?)"

In part, this may relate to their poor performance in the experiment but these comments were reinforced later with a query about the role of such social network software being of use to researchers and/or research projects.

Participants also highlighted a number of technical barriers to sharing that these particular applications contained. This included concerns about the amount of time it took to set-up the software, in situations where "you REALLY [sic] need to share GI" and that the applications were cumbersome given issues of:

- a complicated interface
- the need to add friends
- the "massive search" that would be needed for sharing data in a real setting

Specifically, a weakness in these technologies was identified in terms of communicating and negotiating sharing with other users:

> "However, you need to tell the other person when the files are available. This means prior communication"

In addition, one participant noted that within Exaroom, a registered user could browse a friend's list of friends, and wondered if they could look at the contents of those secondary-friends' shared folders. The group was a closed one (with the exception of 'Major Tom') and adding less well-known members to the group could have further impacted on some users' trust, which will emerge as a theme through this report. Indeed, concerns were specifically raised about privacy issues, where it was noted that AllPeers was less technically invasive (as it is a Firefox extension), although it should be noted that at least one participant's level of concern was not as strong as the statement may seem:

> "Exaroom: who the hell is Major Tom?... I do not want anything from Major Tom running on my systray"

In contrast, another participant noted that, from a mainly technical perspective, "Exaroom is a bit better as it is not client based", as AllPeers had not been able to support the participant working from both home and work. The outcome of this included the creation of multiple online personae or participants simply being 'locked out' of their preferred system. This can be contrasted with a couple of comments at the workshop involving the idea of being able to use technology to share data with yourself (through a variety of

methods). This is something that Exaroom encourages users to do but was not explored further.

For two participants, at least, there were acknowledgements that such products and activities are still very new. Comments included the fact that, in terms of their potential role, few examples have been fully established as yet and that although there was "probably a big role [there were]… many problems… apparent from the many e-mails last week" during the experiment.

3.2.2 Potential Roles of the new informal methods

Some (typically technical) participants saw potential in both AllPeers and Exaroom. A mixed view noted:

> "I would like to use them to replace my normal methods, but find them too cumbersome."

This statement, perhaps, indicated the weakest case for their adoption. However, another felt that the packages were "similar [to my current methods] but enhanced", with similarly positive comments highlighting the underlying technologies place in sharing large GI datasets:

> "Peer-to-peer solves the problem of large files (for email) and [the lack of]... immediacy for CDROM… AllPeers within PortableFirefox is brilliantly portable for any PC, non-invasive and secure."

In part, this was the expected outcome of the experiment and why the two packages involved were used, as data volume and speed of sharing is a concern of the community who are limited by e-mail or slowed down by posted media[11]. In contrast, the strongest positive themes identified by the participants can be termed 'community', 'interaction/communication' and 'collaboration'.

Firstly, one participant noted that AllPeers and Exaroom exposed a greater sense of 'community' and user-identity compared to other approaches, noting the "enhanced... sense of community [and]… visibility of users". User-identity is a long-discussed topic for online environments and, in at least one instance during this experiment, one user was not readily accepted as a 'friend' because they had chosen a user-name not recognised by the others. When the automated request came they were rejected by the group- highlighting a willingness of participants to exclude as well as include in their community- driven by notions of privacy, security and trust, discussed further below.

[11] It should be noted, however, that some instances may remain where it is more suitable to receive data in this way, such as the costs for the Ordnance Survey to provide its most detailed dataset, MasterMap, to the UK HE and FE communities through additional server capacity, alongside the current service being piloted by EDINA for a launch in September 2007.

This notion of 'community' is, arguably, stronger for GIS-users because of the nature of geospatial data and the types of analysis it is used for. Although there is not the opportunity to explore this in detail, examples include:

- the frequent activity of joining attribute information to another's geographical elements (obtained from official sources or other researchers' datasets- hence the tasks in the experiment)
- contrasting the data collected from a study against another's representation
- combining datasets on the same topic from several sources to extend the geographical extent of a researcher's dataset.

Secondly, linked to a network of users as a community, notions of 'interaction/communication' also emerged. As well as noting that the "… built-in IM client [was] very handy", participants noted the immediacy of sharing on offer when friends had already been added:

"… the communication tool linked to sharing seems quite useful- [one] can do things there & then".

Specifically, interaction was noted in two cases, where the tools were seen as "…more interactive/in your face" and that they ".. enable[d] data sharing across multiple users in an interactive manner". It is interesting to note that the former comment could be taken to have a number of meanings, including the immediacy of the interaction but also the potential intrusive or disruptive nature of (at least) the IM application.

A more passive form of interaction was discussed by two participants who felt that informal repositories could be "… useful to browse other people's data" and could potentially reduce the time spent tackling "…normal data requests from students". The latter example led to some discussion at the workshop about sharing for this purpose and the creation of links akin to Frequently Asked Questions (FAQs) in the form of 'Frequently Asked Datasets' (FADs). Both comments indicate varying degrees of interactivity when sharing data:

- The normally asynchronous interaction where the 'supplier' puts information in a shared space and a 'browser' looks for the data, possibly without prompting
- The more responsive case where the supplier reacts to requests from browsers and places information in a shared space so that subsequent browsers may not need to approach the supplier.

Such activities are worthy of further exploration but are beyond the scope of this study, except to note that participants were later to reflect on the use of a formal repository to support datasets for teaching. A final area worth reporting was the impact that this part of the research had on at least one participant's views of sharing, in general, noting that they

"Could see this as encouraging me to share more frequently and differently

(i.e. photos, video as opposed to just data)"

This links to AllPeers and Exaroom being developed as social network tools but the comment also indicates the important idea that, by participating, their view has changed and through 'action research', where participants' views are developed through the strong educational elements in the research activity. This varied from understanding more about the social issues involved in sharing GI to how to install peer-to-peer software on managed systems.

Thirdly, building on both these notions of community and interaction/communication is the more holistic area of 'collaboration', both in general and for research. Although some technical weaknesses have been highlighted, the informal repositories were clearly identified as "collaborative methods". Their role for group-work was also emphasised in several instances. Examples indicated by the participants included:

- facilities to support staff in "team teaching" activities, where the tools could be used in the context of workshops or practicals
- facilities for student groups to aid project work
- "staff-student data exchange projects?", as a possible aid to developing activities such as online training activities;
- research projects which needed "several people to work on datasets together (?)";
- "They might have value whilst working on a particular project with, for example, people at a university with whom you have no shared resources."

These 'problems' and potential applications reflect the participants' approach to how they value such technologies and wanting to see them used in certain circumstances and to varying degrees. They have also clearly uncovered a mix of both social and technical issues.

3.2.3 Outlining the participants' technical and social issues

By contrasting their current activities and the new methods, participants were asked to reflect on any technical or social/organisational issues relating to either the experiment or the methods used. This helps to expose views which inform some of the practice of geospatial data-sharing, in the experiment or in general.

Beginning with technical issues relating to the experiment, it is clear that even well-supported and directed 'simple' information-sharing exercises need a lot of negotiation and time to reach a realistic sharing scenario. Part of this relates to the participants' familiarity with AllPeers and Exaroom:

"… [I] probably would do better if more familiar with them"

"… but I'm still not that good a user yet…and have confused which package does what"

One participant noted it was "difficult to compare two different softwares at once". Although direct comparison of the two applications was only a partial concern of the experiment, participants' work practices and installation/configuration issues of the software were particularly noted as problems:

> "I was between working at home and the office with no Exaroom and/or GIS!"
>
> "Trying to access the systems on different machines was a problem in AllPeers… I couldn't access AllPeers on my laptop at home after creating my account on the PC in work"

Other issues with the experiment related to not understanding their roles in the experiment:

> "I mis-interpreted the "round-robin" (no pun intended!)…I anticipated the prior person in the chain would contact me to let me know the data was ready."

The participant making these comments was quite late on in the failed 'chain'. Some negotiation took place by e-mail but participants were allowed to freely choose the way in which they would share data and interact with the rest of the group. Allowing this free choice was an attempt to create a 'realistic' setting to help expose technical and social issues, as too much guidance would have protected participants from these everyday but important concerns.

Specific problems with the technology were also highlighted. Although it took only minutes to install and register for AllPeers and Exaroom, some participants felt that they took "… too long to setup, [with] too many clicks". They also noted problems logging on to the appropriate websites and that the "continuous login" needed as part of the experiment was not always possible, particularly for those who "… needed to be logged on (as administrator) all the time", which also raised issues for another participant querying if, under such circumstances, data would be "… accessible 24/7?". For those with administered systems this presented further problems including several copies of Exaroom software appearing to run simultaneously and causing concern. This variation in experience and ease of installation and management of the informal repository software certainly impacted on how they would subsequently value such technology as a means to share GI.

Although only noted in passing, the size of datasets was also mentioned as a potential concern. The second part of the experiment was supposed to take this issue further but this was not achieved. If anything, this topic seemed of less concern to the group, partly because of a general opinion that the technology would become available to allow sharing to take place with current

'normal' volumes of geospatial data and that they could cope with little inconvenience using their existing methods.

Social issues with direct links to technical problems were also highlighted by the participants. In at least one case, "IT paranoia" was highlighted as an issue that information-sharing seemed to be uncovering. Participants were also concerned with potential privacy and/or security issues, in terms of topics such as "data privacy?", "Holes in the Firewall" and a participants university not being "…in favour of P2P software in general". In both the case of 'IT paranoia' and privacy/security issues, a general view of 'trust' also emerges, often cited as a main concern of those investigating information-sharing. Trust in this case appeared to relate to:

- who the software was being provided by and
- what the applications could be doing to their systems

In addition, a participant wondered if there was a "… willingness to install Firefox (or portable apps) by recipients", again highlighting the idea that information-sharing is dependent on the ability/capacity and infrastructure of others for the process to be accomplished that perhaps impacting on trust in terms of 'reliance'. The trust issue also emerged in terms of online personae. Although everyone had e-mail addresses from all participants, one participant felt it was "… difficult to find people if they hadn't let me know their username". User-interaction also seemed to be a trust issue, such as the requirement to have friends 'sign-up' to have them validated as a user to instances where one participant noted that

> "We're not asking each other for credentials really… You have to establish that you trust/know who a friend is".

It should again be noted that everyone was working in the context of a research experiment and that previously established relationships (that existed in a number of groupings) would play a role in establishing trust between those involved and that this could change as less well-known people could have access to their systems. In addition, interaction and trust in the systems also emerged, where a participant noted, "I couldn't tell when or if the person I'd shared with had received the data". The large-dataset part of the experiment involved sharers informing the facilitator and the recipient when the data had been shared and not necessarily the group. The lack of this interaction and acknowledgement seems to be a weakness in data 'push' systems. In addition, this problem in the experiment has helped to uncover issues surrounding both the extent to which those sharing information trust the system to deliver information effectively and the need to interact for this purpose, in either synchronous or asynchronous situations.

Trust also emerged in terms of data quality, partly in terms of knowing the quality of the information being made available but also from those doing the sharing: "[I am a] little reluctant to share in case data is inaccurate … but [I] also want it out there so it can be checked".

To some degree, the latter part of this comment relates to valuing data-sharing for the 'greater good' of the community but that its perceived quality would possibly prevent the way in which it was shared (in terms of, for example, completeness or currency). This trust issue can also be seen from a desire for some participants to "… want to make sure I am acknowledged", linked to ideas of data ownership. Specifically, one participant showed a concern for

"Issues of trust (relating to data and how it could be used)"

Such comments illustrate the relevance of 'trust' to data-sharing but it should also be acknowledged that other notions of trust will exist and require further exploration.

In addition, the participants contrasted AllPeers and Exaroom with other technologies, such as the "alternative… [of a] simple group webspace" and the "GRADE [demonstrator] Repository?", itself. One participant went further, asking:

"Do my 'friends' want a shapefile?... Do we really need social networks to share effectively GI?"

The context of this comment included ideas about Exaroom and AllPeers being services designed to share information with (real) friends, with particular features that could, to some degree, be used in a research context. This calls into question a possible separation between work-based activities and instances when a person wants to interact with others online socially, with different technologies being adopted to keep the two apart. As such, just because a technology can be used to share geospatial data, does not mean it will be readily adopted by the community. The second comment extends this idea further, specifically questioning the role that online social 'spaces' can have to aid geospatial information-sharing, or possibly negotiate access to such datasets. Some of the ideas emerging from participants' shared experience from the experiment can be contrasted to their general view of sharing GI, both before and after their involvement.

3.3 Addressing notions of sharing

Participants were also asked to outline keywords/phrases demonstrating how they viewed information-sharing in the specific context of GI. Again the responses were grouped into more technical and social issues. It should be noted that all participants held mixed views.

3.3.1 Before the experiment

Opinions about data-sharing relating to technical issues focussed around both the technology involved and the actual data being shared. Participants felt

that the "Technology [was] ready now" but that, in the case of the experiment at least, concerns would be raised about "… which is the current version?" of the software being used. Overall, current sharing in the community was through e-mail or CD-ROM, and that the idea of peer-to-peer tools should really be contrasted with: "… 3 letters: f t p [File Transfer Protocol]". Certainly, this is the most established technology for sharing files on the Internet (for the technically capable) but it was also noted in the workshop's discussion that not all universities were allowed access to normal FTP services. At least one participant felt that "people are often unwilling to do anything else", in terms of considering the adoption of alternative technologies. Another noted being "… apprehensive of data corruption" that many existing tools offered (particularly e-mail), linking concerns of data-sharing between the technology and the actual data involved.

Limitations to sharing were noted in terms of only "…basic geodatasets", possibly in terms of technical restrictions (such as file-size) but also in terms of what is referred to as 'framework' or 'core' data. In at least one case, Open Geospatial Consortium (OGC)[12] specifications were noted as a feature of sharing GI but this topic was not raised by many people participants, which may be important given the OGC's mission. Certainly, throughout the workshop this important organisation for the sharing of geospatial data was only mentioned a handful of times and awareness of their digital rights management activity for GI, GeoDRM, was not raised.

Several social issues emerged through this part of the project too. Participants noted positive statements about sharing, believing that "GIS inherently needs this" and that it is a "sensible" thing to do. They also noted external drivers for sharing coming from wider infrastructures such as e-(social) science and "collaborative research", and that sharing "always sounds good in principle". The last comment illustrates the variation that can occur between what the community wants to happen but that, in practice, barriers may be in place. In particular the "willingness" to share was noted by some participants, where comments reflected both being personally in favour or considering whether "…people are unwilling to share data"? In addition, this topic was contrasted with the "Naivety/Ignorance" that was felt to be present in the community relating to sharing GI. Certainly, some participants felt that data-sharing was "Limited", "Uncommon" and, perhaps, "Untested?" before they took part in the study. In turn, this may relate to other queries about the wide-spread practice of sharing or the availability of appropriate tools that could be used. In addition, ideas of data-sharing being "local [but]… Loose", "monitored" and "time-constrained", uncover some of the difficulties in investigating the topic, as sharing may be somewhat 'hidden'. The notion that, overall, sharing was "difficult" was also attributed to license "restrictions"/"issues", linked to the digital rights issues, also of concern to the GRADE project.

[12] The OGC is an organisation established to help foster the sharing of geospatial data through the specification of standards, including Geography Mark-up Language (GML) and web services architecture for GI. See http://www.opengeospatial.org

3.3.2 Since the experiment

In response to outlining their attitudes to sharing GI since participating in the experiment, participants noted fewer technical issues, although some felt that the "Technology [was] an effort" and that users "… need to be net savvy". Social issues emerged relating to the "slow adoption" of sharing practices and that sharing was, again, seen as "Occasional" with "… no critical mass as yet". One participant felt that adopting sharing practices would involve "too much of a cultural change" and raised issues of "control" that users may have of who uses their data, seen through the example of "acknowledgement" or citation.

In a general sense, the experiment led one participant to believe that the process of information-sharing was "messy", although another felt that the informal methods were excessive- like using a 'sledgehammer to crack nuts'. This was linked to a discussion of other "… media sharing systems... [where] kids can use those with no problems!" and that if the tools used in the experiment were being made available for research, different groups of the research community could possibly adopt different approaches, linked to their possible small-/medium-group applications (noted above). More positive statements were also made, where sharing was "sensible...possible… tested… successful" and "reliable… seems sure", which can be contrasted with the perception before the experiment of being somewhat "untested".

The 'action research' element of the project was again highlighted by this and comments relating to data ownership:

> "[I have] more uncertainty (in terms of whether I can share some datasets)"
>
> "[I am] more aware of the problems of identifying which data is entirely 'mine' to share – much of it was created with colleagues or trainees"

This appears to be a challenge to resolve in terms of digital rights but one of these participants also noted being "… more interested [in data-sharing] (certainly for serving up student data sets)". This highlights an important dynamic of the process of sharing. It is educational and, although problems exist, there will often remain a willingness to resolve them where, for example:

- advantages to sharing can be maximised
- data policy issues readily permit

Certainly, this participant also felt that information-sharing

> "… needs explicit inclusion in project proposals… especially for research/consultancy".

This introduces two views as, firstly, appropriate technology may be needed to link between data produced in projects with external partners and that,

secondly, the need to acknowledge the influence that research governance/ethics can have on negotiating future use and access to data. This is a particular concern for social dataset users, in a discipline which relies heavily on re-use but is curtailed by changes in research policy that do not reflect, for example, social science research practice.[13]

In another area, one of the participants felt that sharing was "clustered", perhaps relating to the small communities that may share data, both in terms of the aspatial 'communities of interest' and potentially the geographical communities of the "local" sharing and the variation in methods from frequent/local to infrequent/'global' sharing activities (somewhat illustrated in section 3.1.1, above). As such, the clustering of information-sharing can be considered in several senses, especially in terms of who is involved, the topic of the dataset and the location of both the data and the participants involved. Lastly, two comments that should be highlighted in this section relate to those who chose to mention the GRADE demonstrator repository:

> "The GRADE [demonstrator] repository is needed and is the way to go"
>
> "[It]… is easy, well structured, straight to the point… [with a] good background (Edina)"

These comments are a result of contrasting the informal methods with uploading datasets to the GRADE demonstrator repository. The last comment relating to a "good background", again highlights notions of 'trust', where the management of the repository by the service who provides (official) access to datasets to UK academic GIS communities is a notable feature. Such contrasts are developed below relating to the SWOT analysis of both the informal methods used in the experiment and the GRADE repository (see Section 4).

3.4 What would prevent you sharing your geospatial data?

Material from a survey from the GRADE project was used to explore participants' opinions of barriers to sharing GI in two small groups. The discussion outlines the participants' specific comments relating to digital rights, metadata, IPR, access to a repository, issues of trust, data misuse, limited content of a resource, technical ability and if information is available online.

3.4.1 Digital Rights

The first topic involved a discussion of "Issues relating to varying or a lack of Digital Rights", an area of interest for the GRADE project. As well as noting a

[13] For example, governance on research with human participants has not separated medical and social science practices- with the principle of 'informed consent' difficult/impossible to obtain in instances of data re-use, impacting highly on any attempt to create new GI from otherwise non-spatial data, something currently being debated for all social science research datasets (see Dingwall, 2006).

general "concern" for the topic, two participants felt that there may be cases where this would be "data dependent", with another stating:

"Licensing may well be an absolute barrier to sharing"

Two others felt that there was a need for digital rights to be "specified carefully" and a general need for the community "…to be aware of digital rights". Although one member of this group also raised a query if "… digital rights have a specific meaning?", partly indicating varying awareness of the issue but also how this could be specified for the specific needs of the GISc community (perhaps through OGC GeoDRM activity). This comment may also have related to the idea of one's rights in a digital context such as the implications of the Data Protection Act, although this was never mentioned by name. Another participant stated:

"If I knew it was… not a digital right I wouldn't [share], If I couldn't work out if it was a digital right I wouldn't [share]…"

This comment led to a further query if there was any "liability to authenticate other users", demonstrating the level of concern participants can have with digital rights and the dangers involved in breaching them. Another asked "How could we set this up?", referring to the management of digital rights but, perhaps, showing a limited awareness of current processes or if these are fit-for-purpose, given the inherent mixed authorship and IPR issues present in 'GI'.[14] This position is further complicated by another comment that mixed digital rights may "… cause problems for those without certain privileges- [for example, those] outside academia". It should be reiterated that the GRADE project is focussing on the scoping of activities for the academic community but participants were keen to highlight data-sharing cases with non-academic researchers. Group 2 participants also raised additional concerns about "copyright", an area that would stimulate potential owners of data to consider the management of their digital rights but also something which would apply to 'traditional' geospatial data such as paper maps.

3.4.2 Lack of Metadata

The second topic involved considering "If there was a lack of metadata to go with my dataset". The two groups viewed this topic from different perspectives, with Group 1 feeling this is a significant barrier and Group 2 noting easily implemented alternatives. Although this was acknowledged as "important", limited metadata also prompted comments about other solutions.

Group 1 participants noted that a lack of metadata would lead to "others not understanding what the data is" and that "This would bother me if I wasn't sure of data source/quality etc.", as well as noting from the producers

[14] It can be argued that GI has a 'special case' in terms of authorship given the reliance on mixing or re-interpreting data to make new GI, as noted above.

perspective, "Yes, I would be hesitant in putting such a dataset out…". A clear example of this issue comes from Group 2:

> "[This is] Quite important [at]… a technical level- sometimes data is obvious… [but] what about variable headings ['labels']?"

In this case, the participant noted obtaining a number of tabular datasets where variables were not clear and where moving data between various packages results in a loss of some of this information; something which was seen as problematic both for sharing with others but also for the long-term use of data by the same person. Another Group 1 participant considered the effort involved in metadata creation, suggesting:

> "I suspect that lack of time will hold up metadata creation and this [would] delay/prevent sharing"

In addition, a Group 2 participant also felt that the lack of "time to document [data] properly" would create further problems, with the group considering this as an important issue relating to barriers to sharing and the development of sufficiently-detailed metadata.

Several participants considered alternatives if metadata was lacking, with a Group 1 participant also touching on the concept of effort and that they "would not upload [data lacking metadata,] although there is a purpose built service that automatically codes metadata". Similarly, another Group 1 participant noted the issue of who the data could be shared with, noting that "… if it is 1 to 1 sharing in a controlled environment, informal comment could be added". This was seen as something that could take place in the context of small group activities (noted above) but also illustrates a desire for more information about datasets beyond metadata. Participants from Group 2 felt that "There needs to be a basic minimum [in terms of metadata], but given that it wouldn't stop me [sharing]" and that this may not be a barrier "… as I could probably write my own metadata and flag up unknowns for users".

Partly, these issues also tacitly raise the area of metadata standards, either through the effort of completing the current desired information or in terms of what metadata should be specified when data could be shared but may not be already present. To some extent it also raises issues of 'collaboration' to help refine datasets through use and re-use, something that has limited consideration at present and requires a theoretical approach beyond the scope of this exercise.

3.4.3 Unprotected IPR and Research Interests

The third topic involved considering "If there was fear of lack of protection over your own intellectual property rights/research interests". Group 1 started to discuss this on the workshop forms, feeling this was a real concern to furthering information-sharing in research contexts. As well as noting that

they "Wouldn't share in these circumstances", they also felt there was a need for "some method of protection/acknowledgement of research/work on/with data" such as through "citation" and that "can protect IPR", particularly "if [it was a] really genuine IPR interest". In addition, Group 2 considered the more general issue of "permission" to use data. This concern was illustrated by the comment:

"[I]… need to know who is using [my data] & for what, & in what way my data is being modified".

Similarly, there was a desire for the "creation of a reference system for GI datasets", in the sense of citation. There is a need to explore what the drivers of this "need to know" are, as a question remains about the underlying issues leading to this position and what the community's sense of risk to sharing would be if this was not in place.

In contrast, one member of Group 2 felt that a lack of IPR protection was simply not a barrier to them sharing geospatial data, except in the case where universities were concerned about researchers 'giving away' valuable ideas or other resources where charges could, otherwise, be required. There was not enough time to discuss this further but there seems to be a need to reflect on the apparent eagerness in some arenas to have universities working in a more business-like mode and what impact this may have on data repositories, either (inter-)institutional, national, formal or informal.

Others felt that IPR issues were "quite important" and that they would have to "think twice" before sharing. Another participant suggested a solution where details would be needed "about required credit [which] … should be included with metadata". Certainly, some current metadata packages include licensing and sharing conditions in some data-fields but there is no widespread culture relating to the need to cite dataset authors with the exception of copyrighted information available from official data providers (such as the Ordnance Surveys[15], the British Geological Survey (BGS) or the Office for National Statistics (ONS)).

3.4.4 Access to a reliable repository store

The fourth topic involved considering the impact on sharing "If there was not access to a reliable repository store". Mixing the views of both groups, the participants viewed this in terms of the current situation, the degree to which this was a barrier and some solutions. From the outset, a Group 1 participant also chose to discuss the wording of the topic, asking "What is 'reliable'?" and if such a service would be free or "subscription-based?", thus demonstrating the impact on resources that providing such facilities may have. Participants felt that this "Could impact- although [there is] not one currently"[16]. In

[15] i.e. for both Ordnance Survey (GB) and Ordnance Survey Northern Ireland (OSNI)

[16] (although examples exist such as the Natural and Environmental Research Council's (NERC), noted elsewhere by a participant)

contrast, a participant in the other group felt that the lack of access to a reliable repository "obviously would 'prevent' widespread sharing", valuing such resources highly. In comparison, three participants from across the two groups felt that this was less of a challenging issue, as they "… would just share it [data] with relevant people". Others felt that they would

> "… still share, but not as much as if it was reliable (such as the GRADE [demonstrator repository])" or
>
> "… still share, but a repository is preferable".

Highlighting the GRADE demonstrator repository is notable but should be considered alongside the current adoption of other technologies such as using USB storage devices or possibly even using "…my own network and share data in it" with specific people. As such, a national geospatial data repository is a resource that has value for a wide base of data consumers and the purposes of sharing should be considered, contrasting file transfer for back-up against data being shared for analysis and the outputs of completed projects towards long-term curation.

3.4.5 A lack of trust in others

The fifth topic involved the challenging subject of "a lack of trust in others" and led to varying views in relation to the topic being "Fairly important" to others that felt it was "Irrelevant", although this may be a reflection on their own approach to trusting others rather than the status of the issue for the GISc community as a whole. Trust was considered in terms of elements such as acknowledgement, authentication, data accuracy and who they were sharing with. Most felt that they would trust fellow GIS-users but under certain conditions, noting that they were "Not worried- unless unsure over licensing issues" or only those that "were signed up [i.e. authenticated]".

One participant felt that there would be greater "trust if acknowledgement could be in place", with others noting that "metadata and authentication" and the "need [for] some standards regarding metadata" possibly reduced risks which were hoped to be "eventually… normal". Two participants went further to suggest that:

> "a central repository should help to address this"
>
> "Digimap restriction on [the] GRADE [demonstrator repository] is useful to mitigate 'risk' …"

Again, this highlights participants' views of a GRADE-type repository could have in tackling the perceived issue of, for example, trust and that the current Digimap authentication process aided this. Others were more concerned, noting a "…need to be aware of what the community is that I am sharing my data with" and the impact of "other/wider access- [, where trust is] more

relevant". This issue of 'who' could be involved was also seen to depend "… on the dataset but I would be cautious", whilst asking "How do you prevent people from uploading incorrect data?". This last comment can be compared to another Group 1 participant who felt they:

> "Might not put out all my data, [there are]… possibly 2 levels of datasets, [with] one that can be shared".

This may, in some senses, contribute to a perception that data may be "inaccurate" just because it is incomplete.

3.4.6 Misuse of my data by others

The sixth topic involved considering the "misuse of my data by others"- a concern linked to trust, as seen in the previous comments. Participants across the two groups considered this topic in terms of its overall importance, users' responsibilities, the nature of the data and chose to suggest solutions to misuse too. As well as one Group 2 participant feeling this was "irrelevant", others noted that they "Don't [typically] fear this". Many participants felt that this could be an issue if people were not familiar with GIS-related research. It was felt that a data producer "… can't be responsible for other peoples' research", unless they were being trained by the producer, as misuse problems would need to be explained. In contrast, another wondered "Do I have any liability?" in the datasets created and then shared.

In one case, causes of misuse were attributed to "data versioning", where someone may not use the appropriate dataset. For one participant they felt "it would depend on the type of data", contrasting between the less problematic situation with OS data "...but if I derived some data it could be interpreted wrong". Several solutions were also mentioned relating to the need for "better control of data" through "… different [levels of] access", including where "Some users can edit, others cannot". Others hoped that "… there would be appropriate/sufficient measures in place" to prevent misuse. Metadata was again highlighted as a core feature to tackle this, something which may require more widespread training in the GISc community.

One participant went further into this policy arena by suggesting that "a system [should be in place] where abuse is reported". Such 'policing' is likely to introduce a new set of 'trust' (and other cultural/social) issues, as well as relying on a degree of transparency for a self-policing system from the grassroots-/user- perspective against a more top-down 'authority' with, possibly, punitive powers. In addition to these concerns, two participants considered technical approaches to aid detection of misuse:

> "[We could have] Electronic copyright embedded in data"
>
> "[I would like access to] log files/update on who's doing what with my data…"

The first relates to discussions of technologies such as 'digital watermarking' or the inclusion of minor errors in datasets so that use without permission can be detected. The second comment may seem to involve a degree of concern about how data is being applied but this topic also offers great opportunities in the context of *in silico* research[17] through:

- Generating an interest in datasets beyond their initial application
- Increasing interaction/collaboration between dataset users, possibly mitigating misapplication through (online/public) discussion
- Monitoring the impact that a particular dataset is having for other users
- Providing a context that helps to support the idea of citation of datasets
- A starting-point for more social tools[18] that current online/collaborative research activities such as e-(social) science could grow to depend on.

3.4.7 A lack of interesting datasets and starting from scratch

In terms of their seventh topic, participants were asked to jointly consider "(a) a lack of interesting datasets to find, and (b) that it is easier to start again from scratch". Some saw this as a "major issue!", although two participants felt this topic was not applicable to them. It was noted that "It is always useful to find ready-made datasets and adapt them...", providing that data "… was easy to find [and that] it would [overall] save time". Specifically, two Group 1 participants felt that:

> "… sometimes the time taken to adapt and modify existing datasets has sometimes meant that I've preferred to create my own dataset from scratch…"
>
> "Creating your own dataset is useful [but] time consuming but at least they are reliable."

Reliability also emerged through another Group 1 participant's comments, where they believed their colleagues would say "'I'll create my data as there's nothing good enough'". In contrast to this, a Group 2 participant felt that "It might be easier [to create your own data] if you don't understand a dataset…", suggesting that this would also rely on "good metadata".

In contrast, two participants from different groups noted that they "…would still share", possibly because "… I had some [data] I was proud of having created" but that they, similarly, "might have no choice" in available datasets, thus leading to them having to start anew. These ideas express a willingness to share even if others are not participating to the same degree and see starting afresh on analysis as a potentially weak approach.

[17] A concept where all research activity takes place online including the storage and analysis of datasets, the sharing of all outputs through linked resources and the collaborative tools needed to foster such work.
[18] (such as Amazon's 'recommender systems')

3.4.8 My technical ability and public access

The eighth topic involved considering another set of joint topics in terms of "my own technical ability and concerns over public access", which participants mainly viewed together. This approach and the greater separation between the topics led most participants to not be concerned with this issue, particularly as technical ability "… or lack of it can be overcome". This also indicates the level of comfort the group had with their ability to use technology in information-sharing contexts. The topic also raised access/control/usage issues for one participant, with another feeling that something akin to the GRADE demonstrator repository could address in terms of access. Another felt that they wanted public access to data "… but the uni [university] may stop me", again linking a notion of access to barriers present in Digital Rights, noted above. To some extent, participants saw their technical capability as less of a barrier than problems associated with wider access to data.

3.4.9 No Internet access

In the last of the topics relating to barriers to sharing, participants were asked to consider "If I did not have access to the Internet/web sources". Group 2 focussed on the likelihood of this, noting this is "Not ever a case" and that it "… shouldn't apply in the UK", with an exception of fieldwork and the use of mobile devices to access data ("[It] depends [if I am]… stuck up a hill!"). However, it was acknowledged that, nowadays, the lack of Internet access would prove to be a barrier in the ability to readily share data. Group 1 considered this more from the impact that loss of connection would have on their activities, with solutions involving the use of the offline media mentioned several times above. One participant considered another approach drawing on available offline possibilities, where someone "could create a service like Netflix", a postal service that allows subscribers to rent films on DVD. However, one Group 1 participant also noted that the loss of connections would make it "… difficult to find fresh data", indicating the extent to which this practice of online data-retrieval has become commonplace in the community. As well as these varying problems to data-sharing the participants were then asked to consider a range of topics that could ease or support data-sharing.

3.5 What things would make it easier for you to share geospatial data?

In contrast to the barriers to sharing, participants were asked to respond to possible solutions noted in the previous study, with both groups only making short statements and Group 1 deciding to report the group's collective opinion.

3.5.1 National geospatial repository

Both groups simply felt that "Having a national geospatial repository holding for the UK" was a "good idea" and that "This maybe has a strong component

to wider infrastructures", linking to notions of the relationship of a repository like GRADE to National Spatial Data Infrastructures (NSDIs, discussed below). Another Group 2 participant noted being in favour of such a resource but queried "…where do you draw [the] boundary and what about ~~incomplete~~ [sic] variations in coverage?". This raises two different issues.

Firstly, what should be the geographical extent of a national geospatial data repository? For example, there are two Ordnance Survey's in the UK- should the repository contain data from both, and is its function to hold data from the UK and/or GI created by researchers based in the UK?[19] This also raises questions about the role of regional and institutional repositories, in terms of being best placed to facilitate the accuracy of this (typically local) place-based information but, in addition, not necessarily being best placed in terms of expertise to look after geospatial data appropriately.

Secondly, a single centre for data covering the UK may be reliant on the information generated by many regional centres which, due to the research interests of staff at these centres, may not produce similar datasets to build towards a 'patch-worked' national picture. Examples include variation in access to crime data between neighbouring regions with different police forces, as well as difficulties associated with defining and recording any variables that could, in practice, be compared.

3.5.2 A central portal for GI

Both groups agreed that "Having one central portal that finds geospatial datasets" was a "good idea". Informal discussion at the workshop suggested making stronger links between the metadata and information service, Go-Geo![20], and the GRADE demonstrator repository. Participants considering this a useful combination of resources, where data could be browsed and discovered through the Go-Geo! catalogue. Certainly, an approach involving Go-Geo! would help support metadata standardisation and also the opportunity to link to other types of information of interest to the UK GISc community such as events' calendars and training materials.

3.5.3 Reducing license restrictions

In terms of "Having less restrictive license agreements for academia", Group 1 noted "with caution" that it was a "Good idea". Group 2 equally suggested that it was a "nice idea" but that it would not happen, as it was "unrealistic" and difficult to implement as it "ties in with the whole O.S. funding issue". There was also some informal discussion relating to licenses actually helping to protect research interests and the quality of data but there was not an opportunity to consider this further.

[19] This idea has been raised in US contexts for some time in terms of distributed geospatial libraries (see Commission on Geosciences, Environment and Resources, 1999)
[20] see http://www.gogeo.ac.uk

3.5.4 No license restrictions

Continuing this theme, the groups were also asked to consider "Having no license agreements". Group 1 felt this was possible but that there was "… some argument [to maintain restrictions] for certain data". Group 2 saw this as much less likely than the previous topic: "No, as this wouldn't work". Another Group 2 participant suggested that:

> "[I] Do not think this would help- licenses are there [for] a reason- [What about] liability?"

This highlights the idea that licenses are a legal device and a means of protection for both the person permitting their data to be licensed and the party receiving that data. It also illustrates some concern that sharing data may involve issues of liability which could involve licensing as part of their protection. In addition, it should be reiterated that this topic was also viewed in terms of metadata creation during other activities at the workshop.

3.5.5 Local/departmental repositories

The topic "Having local/departmental repositories" led to several comments from Group 1. They felt that it was "Not a good idea [for sharing] but a good way of accessing data as a back-up", believing that all departments "do this as a way of sharing data between colleagues". Group 2 noted variation between the two institutions they came from, where two participants noted having a local/departmental resource, and the others feeling these were "Useful… " but that it "… makes sense to move to national [repository]", although no specific reasons were expressed. One Group 2 participant felt that these resources may have a greater role in a "local-to-global information management perspective", given the comments noted above in relation to many local sources contributing to one national resource.

3.5.6 A GIS forum

Group 1 saw "A geography/GIS forum similar to Napster or MySpace" in mainly negative terms, feeling it was "… not a good idea" and "Invasive". They also raised issues of "Access control… [and] Who is accessing the system?", relating to previous concerns noted about control over their data and possible problems with such peer-to-peer technology. One positive element from this type of resource was that it could be "… useful for auditing available data". Undoubtedly, their experience from the experiment has shaped these views. Two Group 2 participants did not hold any views about this topic, whereas one of them noted that:

> "GIS forums exist… [and that] Comments attached to data might be useful (i.e. users' comments)"

Similarly, another Group 2 participant suggested that "Some forum around data/research could be useful". However they noted not being familiar with Napster/MySpace's "community tools"; that there may be issues with "the

number of participants" involved in such a service and how affective it could be, again possibly relating to their experience from the workshop. This introduces more ideas of the discussion of data *in silico* beyond metadata, noted above.

3.5.6 Control over data usage

As already noted, participants from Group 1 had concerns about "Having greater control over who used your data", stating that it was "Important to know who access[es] your files and what they are using them for". Group 2 participants, in contrast felt that this concern was either "Irrelevant" or that, although they had no interest in control, it would be "interesting to know" how their data had been used. The discussion in this context also included research practice and data-sharing issues, where it was felt that sharing should be "embedded into consultancies' research projects", again linking to the need to be clear with research participants about how data could be re-used.

3.5.7 Faster computer speeds

Group 1 took a purely technical perspective of the desire for "Faster computer speeds", noting that it was "Not relevant at the client end [but] may be for [the] server end". Group 2 acknowledged that this was "… only an issue if it was really slow" and when "… dealing with large datasets only". One participant was particularly keen on this topic, given the amount of data-processing involved in their work, which led to the group acknowledging that computing speeds are important but that, currently, the technology is progressing in line with demand and that "This is only a minor issue compared to the others [discussed in this task]".

3.5.8 Listing Informal Sources

When asked about "Having a list of all available sources of informal geospatial data", Group 1 felt that this would "Only [be] useful if linked to data or has more detail about the data" and that a list alone would have limited utility. Group two participants presented several comments including, that it was a "good idea" with the potential to "… tie in with a repository". One participant also wondered if

- it would be possible to find GI in these circumstances and
- if this would only include a list of all "Napster-type things"?

This was partly answered by another participant noting the potential to develop "registries".

> "A registry is a set of special e-services that support organisation and discovery of and access to online data and processing services... Registries help users or application software or other services find or retrieve data or e-services existing anywhere in the distributed computing environment" (McKee and Pichler, 2003; p. 8)

In addition, the participant interested in how to "find GI" also considered that, once any of these datasets were found, should they be moved to the 'safety' of a more "formal location", such as a geospatial data repository? The limited discussion of this OGC-related topic by the participants may indicate the limited understanding of, or opportunities to implement, registries and other e-services, although being able to discover data was valued by both groups. An additional comment was added to this issue by a Group 2 participant, in terms of concerns about the adoption of OGC Standards for services like the GRADE demonstrator repository to increase their interoperability and connection to wider infrastructures. In addition, it should be noted that the hosting of data in stable and 'fixed' online locations is somewhat of a prerequisite for registries to develop.

Throughout the discussion in this section, it can be seen that the participants have had varying opinions of the various components that come together in the context of informal (and formal) repositories to share geospatial data. Such a comparison was also examined directly through SWOT analyses of both the informal methods of AllPeers and Exaroom and the more formal approach of the GRADE demonstrator repository, discussed in the next section.

4 Results (Part 2) SWOT Analyses

4.1 SWOT analyses: Informal Methods

4.1.1 Strengths of informal methods

Apart from one participant who saw limited application of AllPeers and Exaroom for GI, participants' comments relating to strengths fell into several groups. As well as noting low cost as a positive element of the software, their technical component was discussed, with participants highlighting:

- ease of installation
- ease of registration (through one login rather than multiple links)
- "ease of use"
- the user-interface of Exaroom being better than other examples
- "less data corruption" in the sharing process
- that the tools could be "always on"
- that Exaroom was available on any machine

Access to information was also considered, both in terms of the speed or immediacy of access offered alongside the ability to control the access to data for groups or individual 'friends'. Similarly, the notions of the informal repository methods being "good community building tools" that were "collaborative" and "interactive" was specifically noted. This was illustrated further by the role of their "communication tools" which helped to "... negotiate [the] technical aspects of sharing activities" between users. Such

facilities were also noted in terms of their potential to "set-up a project group" and offer an information resource "…if there was a change in staff".

4.1.2 Weaknesses of informal methods

Again, weaknesses were considered in terms of the tools' technical and social aspects. AllPeers and Exaroom were weakened by their technicality in terms of:

- "Too many features not really needed".
- AllPeers being "client based"
- Exaroom being both "…invasive and [requiring] the DotNet extra download"
- "Privacy" and "possible firewall issues"
- A need for administrator login for both installation and use

Participants also considered the appearance of the software:

> "I don't think AllPeers has cracked the user interface as yet. It is only beta software though"

The second part of this comment is significant as nearly all the current applications reviewed in the study are beta products, with their "untested" nature noted as another weakness.

Function also emerged as a weakness through one user's "need to 'chat' to multiple users not just 1-to-1" during the experiment, particularly when negotiating the sharing of data between multiple users.

Although easy to use, participants suggested the informal repository methods were "not conducive to use", being confusing and unfamiliar. Specifically, barriers to sharing present in the technology in more real-life settings included a view that:

- "[The] Computer needs to be on 24-7"
- "[You] need to be logged on"
- "They Require both sharers be pre-registered as friends"
- "Require both sharers be logged in"

As well as these issues of interaction, limited abilities to potentially obtain data were attributed to the lack of "a single 'location'" for the data or their "distributed" nature impacting on finding the information in the first place, something which can be contrasted with the role of a national data repository.

As the informal repository methods "… are not built to share GI" they were seen as having:

- limited utility,
- "not interoperable"
- limited "organisation of data" in these packages

What has not been determined, however, is the extent to which the 'interoperability' is being taken in the sense of, for example, technical or semantic interoperability. Certainly, the semantic element should be of interest, especially as one of these two participants also raised issues of judging the "quality" of any data being shared.

4.1.3 Opportunities of informal methods

Participants consider a handful of opportunities seen through the packages and the experiment. Some of these focussed on data, from the perspective of sharing and policy. In terms of sharing, two topics were noted, providing 'Frequently Asked Datasets' for teaching and file-sharing across multiple users as "data resilience… in case things are deleted by accident" and the long-term maintenance of data through digital curation.

Opportunities, seen in terms of data policy, were highlighted by one participant:

> "There is a real need for data sharing to be embedded within academia [and]… there is a need for 'best practice' on sharing to be developed"

What could be viewed as "best" remains to be seen and further work is needed to develop this view in line with how a repository infrastructure is being developed, promoted, supported and maintained.

A significant feature of the informal repositories opportunities was their potential to "aid collaboration" under five inter-related themes:

1. "Quick negotiation" of access to data "A: 'Have you got hospital locations?' B: 'Yeah, in my Exaroom folder'"
2. Developing "data sharing communities" to expand "the personal network of GI users"
3. Projects where a small group needs to share datasets
4. "Inter- and intra-institute collaboration", noting the need to share data within and between organisations
5. "Working with non-academic partners on projects" and "non-academic access" to datasets

This last topic is beyond the current GRADE project's 'scoping' activity for solely academic users but illustrates a real research-practice issue, where concerns were raised by participants about mutual access between academia and the public sector stakeholders (in particular) when trying to share geospatial data (or even data that could be turned into GI).

4.1.4 Threats to/from informal methods

Again technological and social issues were considered as threats to both the use of informal repository methods and geographical information-sharing, in general. Those participants who were more technically-oriented noted:

- the software was "untested"
- a dependence on "…on 3rd party software"
- that alternatives existed to Exaroom and AllPeers
    - that "personal preferences" would impact on selection of alternatives
    - that better or more embedded examples may exist
    - "…with more GI-related functionality"
- "… other systems [were] not compatible"

This second point was also linked to the limited technical support for such systems and the potentially slow adoption of new versions to meet changes in the surrounding technical/technological landscape.

More social issues considered:

- the limitations of having an infrastructure which "relies on individuals to run and maintain [it]"
- the administrative overhead for valuable data assets could be prohibitive in some local settings
- system security
- "invasive" nature of the systems and "surveillance?" (especially when it offered too much scrutiny of colleagues' work)
- the need to "trust" software providers who could, potentially, have access to the user's data
- the loss of "… control over your data" once it is made accessible

Lastly, three comments related to the way in which such technologies are adopted:

- The "slow take-up" of technology and sharing practices, impacting on the culture of sharing in the academic community.
- "Will the community really grow?"
    - limited number of users
    - limited resources
    - technologies difficulty of use
- The lack of development of best practice, with the current culture and "academics!" identified as a significant threat

These views can be contrasted with the SWOT analysis of the more formal repositories.

4.2 SWOT analyses: the GRADE demonstrator and other geospatial data repositories

4.2.1 Strengths of Geospatial Data Repositories

As with the informal repositories in the study, participants felt that the GRADE demonstrator repository was "quick" and "easy"/"straightforward to use". Particular strengths were highlighted in terms of:

- "reliable" and "robust"

- "permanent" and "dedicated" resource in a "single location" (aiding "collaboration")
- "make[s] data available" as part of an "always-on resource", including, otherwise, "orphan datasets"
- "server based"
- using "single registration"
- the ease of adding and searching for datasets
- "reduced effort" and/or "removal of effort" in creating "shared" datasets
- GI centric / "specifically established to host UK HE [Higher Education] GI"
- Moderated content (i.e. datasets from inside or outside the UK) by a "known/trusted resource" (i.e. administered by EDINA)

This latter point may be of interest to the concerns in wider digital infrastructures such as e-(social) science, where single-point user-authentication is being developed (such as Shibboleth[21]).

4.2.2 Weaknesses of Geospatial Data Repositories

In general, weaknesses in repositories were focussed around technical and data policy issues. A technically-aware participant wondered if the current GRADE repository "…is an OGC compliant W*S [Web Map/Feature/Cover/etc. Server/Service]". This could be viewed in terms of this participant's:

- support for the development of such features
- need to have these things in place to connect to such resources using their technical infrastructure
- need to know if such facilities are in place to develop appropriate resources

Discussion at the workshop exposed this particular topic in terms of Digimap's pilot web service, where data can be accessed directly from desktop applications such as ESRI's ArcGIS. If such developments were not made available the participants felt that repositories could also be "easily inaccessible", preventing ready links between users, their systems, data and outputs towards *in silico* research, also including:

- a lack of "links between submissions" and users not being able to "… sign up to topics/themes"
- "lacks communication/community tools" present in the informal methods

The GRADE demonstrator repository was seen as somewhat "time consuming" to use, with concerns relating to data and metadata:

- "restriction on data types", partly in terms of a couple of development issues relating to loading specific formats during the experiment[22].

[21] see http://shibboleth.internet2.edu
[22] such as ESRI's ArcInfo .e00 files

- Limiting repository content to certain types of 'GI', as GI can be considered to take many forms beyond databases and text files, least of those emerging from considerations of geoinformatics in relation to the semantic web.
- The repository's underlying software "… cannot read GI in native formats" as .zip files need to be submitted (a barrier to some less technically competent users)
- General concerns of the management of "data quality issues", "quality assurance" and "quality control" of data and metadata
- Unknown 'meaning' of datasets, again, relating to the desire for formalised or community-mediated "semantics"
- Problems of "repeated metadata creation"
- The lack of opportunity to "… use previously created metadata" possibly using the metadata already created in ESRI's ArcCatalog product

As well as quality control, participants expressed some concerns (also present in the informal repository methods) in terms of "privacy control" and the fact that users "don't know who is downloading… [their] data" and the well-known issue of varied understandings of license restrictions. It is actually notable that little of the workshop's time involved this latter topic, allowing discussion to focus on other issues and topics found in this report. Certainly, usage of the GRADE demonstrator repository (and a request to upload data to it) presented several of the participants with issues about what data they could share, particularly in terms of finding suitable "derived datasets", another illustration of action research.

4.2.3 Opportunities of Geospatial Data Repositories

As noted above, the use of the GRADE demonstrator repository promoted a lot of positive interest by the participants. From the outset, it was noted that "no-one has figure out the best way to do it", in terms of creating a national geospatial data repository. The creation of such a facility also allows the opportunity to develop:

- the community's "… most common data delivery/sharing method"
- an experimental base for additional applications such as the "implementation of a powerful GI search engine"
- a chance to develop best practice in this arena
- data quality mediation through "Definitive datasets augmenting 'official' representations and descriptions"
- making data accessible to increase quality through re-use (alongside the reporting of errors, omissions and potential application areas)
- aid "collaborative research"/"collaboration", with examples including "project based file share hosting"
- a "Potential major resource for e-science"
- "Integration" [of the GRADE demonstrator repository] with other EDINA GI collections, including the Digimap service
- "links with other metadata catalogues (Go-Geo!)"
- resources such as a "Teaching repository; Research repository"

- Once developed considering the resource as part of a NSDI

These opportunities were conditioned by the idea that a national geospatial data repository for the UK academic community should be "easy to use, comprehensive, relevant [and]… publicised", which can be considered alongside potential threats.

4.2.4 Threats to/from Geospatial Data Repositories

To some extent the participants viewed 'threats' as more powerful weaknesses in the available resource. So that, in one case at least, the limited acceptable "data formats" were cited as something that could jeopardise the wider up-take of the GRADE demonstrator repository. Other threats included:

- potential "underuse" of a resource
- possible concerns about unequal levels of contribution to a resource by some users compared to wider use of that dataset by those who do not contribute as data 'free-loaders'
- licensing and authentication issues,
    - the "need to be certified"
    - the need to "… address community concerns of digital rights"
    - "… licensing issues – perceived or real"
    - "copyright"
    - "…data and software licensing" restrictions
- The sustainability of the resource
    - linked to "funding",
    - "maintenance [costs]"
    - current "Project Shelf Life"
- Metadata
    - Limited coverage
    - the need to adopt "Metadata Standards"
    - the need to utilise metadata already created in software such as ArcCatalog (as noted above)
- Data volumes ("Does the repository have the right tools to cope with increased data volumes?")

In terms of the last point, the suggested opportunities for the GRADE demonstrator repository to act as a test-bed for Geospatial Information Retrieval (IR) may, therefore, gain greater weight.

The wider GISc community's perceptions were also acknowledged as potential threats including:

- That such demonstrator resources may not be seen as "proper" or only as "toy" and not worth participating in if it is not a mature and established resource
- that the resource could be "limited by access rights to academics only"
- the "Core purpose [of repositories] may not reflect research practice in terms of the [current] culture of sharing GI"
- the wider "need [for] a change in attitude" to sharing GI

- 'cultural inertia' issues, including "personal preference" and what is seen as "appropriate" methods to share GI
- Alternative technologies/resources that may exist in competition to the GRADE demonstrator repository ("There are already other geo(spatial) repositories in the UK (e.g. NERC)")

In terms of this last point, what remains to be seen is the extent to which other research councils or services (such as national data archives) foster the development of data repositories; the extent to which they can affectively manage, curate and make accessible geospatial data; and what the implications would be for a tapestry of varying data sources, both in terms of its impact on research activity and missed opportunities of developing a dedicated and trusted resource.

5 Discussion and Conclusions

The sharing of GI through the various technologies outlined in this study requires closer examination and interpretation. This section draws together the evidence from Section 4 under several themes relating to sharing geospatial data. The first part briefly deals with the impact that the study has had on participants, before looking at the participants' views surrounding the sharing of GI in terms of drivers, barriers, problems and possibilities.

5.1 The impact of the study- action research

From a practice-based perspective, it can be seen that the study has developed into a form of action research. Participants found the exercise an educational experience relating to information-sharing and the technologies this can involve, especially as they questioned the role that social network technologies could have as informal repositories for sharing GI. This ranged from disapproval of their use for research purposes, including popular resources such as Napster, to instances where sharing scenarios (such as group activities) could employ Exaroom or AllPeers. Other examples of the outputs of action research included increased uncertainty about what data could be shared (in part relating to ownership issues), where they had not considered these issues before, and that by participating in the study they considered how else they may use such technologies, such as sharing photographs with friends or exploring better mechanisms to share data with students. From a research perspective, the participants have also developed a more reflexive view of information-sharing that has contributed to the apparent honesty and depth of coverage in this report's discussion and, for some, an interest to actively participate in similar/subsequent projects.

5.2 Current Practices of Sharing

Participants illustrated their views about sharing GI in terms of the methods adopted, with whom and for what purposes, whether these involve active or passive sharing scenarios, the purpose of sharing, 'grey-sharing' and the drivers and barriers that may be present in sharing.

5.2.1 Technology:user:data perspectives

The study involved several technologies that could be used to share datasets and related information (such as metadata). Participants currently use several strategies to share GI with a range of people. Common amongst these is the use of USB storage and e-mail for small datasets and burned media for larger ones. For those that are more technically aware (which includes a good proportion of GIS-users), FTP was noted as a method for making data available online, with personal webspaces offering a ready interface for others. In two instances, participants highlighted dedicated online resources to aid sharing (VKP and SCP), with the former geared to supporting project management and collaboration and the latter aiding the secure transfer of data between machines for processing. These are highly purposive approaches and illustrate quite sophisticated means to share data, either alongside other project information or even when just sharing data with oneself.

The purpose of sharing should be recognised, such as work practices including teaching, group work or when users need to share data with themselves, including when moving institution or between work and home machines. Importantly, the GRADE demonstrator repository was emphasised by the participants as a potential method with characteristics that will be explored throughout this discussion and that the informal repository technologies were mainly untested beta versions, which required other established technologies (such as e-mail) to aid their use.

Additional to these main concerns is the idea that sharing takes place at varying geographical scales. To some extent the use of USB is certainly a same-room technology, closely linked to whom that data is being passed to (i.e. where users work in the same place). Additionally, it should be recognised that users are frequently interested in data about 'somewhere' and that this information more often tends to be located within these places. For example, we would expect to find more data about Edinburgh in Edinburgh and that those with an interest in the data from the city are also likely to be based their too (see CGER, 1999), perhaps leading to the widespread adoption of readily accessible methods such as e-mail and USB. This, however, does not rule out another dynamic relating to the nature of the data involved so that, for example, participants also identified datasets that they shared with others overseas, mainly through e-mail and burned media. As such, we should acknowledge that location plays some part in the ways GIS-users choose to share data through varying technologies but that other drivers are in place to encourage the use of a variety of methods and not just

one approach. As such, believing data-sharing to be "clustered", as noted by one participant, may be a notion worth explicating.

As noted above, the datasets themselves are important features in data-sharing. At a simple level, the volume of data present in a dataset can clearly impact on what technologies users employ. However, the content of any given GI dataset and any data policy issues that are attributed to it are important dimensions too. Least of all is the notion that some datasets are fundamental to sharing, recognised as core or framework data in the context of the defined and structured resources, such as SDIs, and the contextual or derived data provided by the Ordnance Survey (GB), for example. For example, a lot of social science GISc could not take place if researchers did not have access to the National Statistics Postcode Directory (NSPD) through EDINA's UKBorders to match attribute data to. In addition, we can readily understand the varying forms of 'geographical information' that can be constructed and that the portability of such data is also related to the systems/software we can use to manipulate that data. For example, datasets created using one piece of software may not readily be used for another. This links to the idea of creating standards to construct datasets, the interoperability of systems and the generation of metadata to aid in the appropriate selection of systems and data, either by the user or increasingly by automated means online.

The specifics of these ideas are being played-out through the part of the GISc community involved in Open Geospatial Consortium (OGC) specifications, where leading manufacturers and other GIS developers are participating in activities to make their systems more interoperable; data-structures are being developed which use a common 'frame' of Geography Mark-up Language (GML) to allow transfer of data across a number of systems; and for web services and, for example, registries to be able to discover and manipulate data online. The importance of this issue in the context of the study is that OGC specifications were raised by some participants in relation to how some information-sharing technologies may not readily support GI (including the interoperability of the GRADE demonstrator repository and that the informal repository methods were not GI-centric).

5.2.2 Active and passive sharing

One of the reasons for adopting certain technologies for sharing should be considered in terms of how the data is being presented and whether this is active or passive sharing; in part defined by purpose, those involved and the volume of any dataset. Data-sharing involves both senders and receivers and such ideas are well accounted for in discussions of most Internet technologies, where it is possible to map-out relationships such as the one-to-one, one-to-many or many-to-many, many-to-one scenarios present in, for example, e-mail, webpages, wikis and online surveys, respectively. What should be acknowledged, particularly in the case of informal repositories, is the 'sender's' perception of the potential 'receiver's' ability to obtain the data. E-mail is a popular method of sharing data (as attachments) due to its

mainstream adoption, through widespread readily accessible technical infrastructure and standards. Senders trust that, by using e-mail to deliver data, the recipient's e-mail server and client will allow this other user to readily download that data. It should also be acknowledged that e-mail is also a communication technology and that should a message fail to be delivered the system notifies the sender, something that was present when negotiating sharing through the informal repositories. User-feedback and knowledge of receipt of a message or data are, therefore, highly-valued by users. Such issues relating to 'trust' will be returned to later in this discussion

Another view of data-sharing includes a more passive way of 'broadcasting' information through webpages, allowing the sender to become a 'disseminator' and for the receiver to become a 'browser'. Such a view links to notions of accessibility of information and the control that sharers can have on their data. At first glance e-mail may seem a more controlled means of sharing but it should be understood that, although a webpage hosting data may appear to be more open, a broadcaster can monitor access of the information, whilst the content of an e-mail can be readily distributed to others without the initial sender taking any further action or having any awareness of where that message has gone.

Such ideas illustrate some of the underlying social issues that shape the participants' opinions and concerns of control of sharing and the technologies adopted in the study. AllPeers is a technology much like e-mail where the sender pushes the data, whereas Exaroom allows a space to be made available for other users to browse to a dataset. Both technologies contain communication tools, something which appears to play a lesser role in the GRADE demonstrator repository but which could impact on how users approach such situations in an active sharing situation. Deposit and extraction of data from any infrastructure should, therefore, be understood to have a social context that may require tools to aid communication, in part depending on whether data is being shared actively or passively for a range of purposes.

5.2.3 Grey-sharing

One of the concerns that the study begins to address are notions surrounding both the datasets credentials and the ways in which the community adopt various approaches in certain scenarios to enable sharing. When the sharing is 'formal' through methods such as the GRADE demonstrator repository the participants noted having to consider carefully what information they were making available. Licensing restrictions and issues of ownership of derived data (or data made for/with other researchers) were considered. This did not appear to be as prevalent a concern with the informal repositories. In part, the design of the study involved sharing neutral data through these informal methods but it is likely that participants would maintain these concerns whether for formal deposit through perceived risks to loosing control of their data. As such, 'grey-sharing' can be considered as instances where formal mechanisms are not employed to share data and alternative technologies are employed to maintain control, particularly for a small group of

users/recipients. This is an area that was difficult to explore with the group, least of all because of the issues of trust and exposure that this would have presented with friends and colleagues.

It is suggested, however, that the group did not value the peer-to-peer examples highly for sharing purposes for widespread consumption of data and that greater risks existed through the perceived security issues present in these technologies. Grey-sharing may exist in the community but it seems to be a practice that is happening on a one-to-one/few basis rather than through developed infrastructures or services, if at all- partly relating to the ownership and control issues that many participants appeared to hold dearly. Few users appear to give away data 'freely'. One possible area for further exploration would be to look for the amount of GI available in global/public peer-to-peer systems (e.g. Kazaa[23]) rather than the informal repository technologies used in the experiments for (typically) confined groups. The remainder of this discussion considers the drivers and barriers to sharing alongside the general problems and possibilities participants considered for information-sharing for both informal and formal repositories.

5.2.4 Drivers to geospatial information-sharing

Perhaps, unsurprisingly, the study has uncovered only a handful of drivers to information-sharing, seen from participants' opinions before and after the exercise. However, such drivers also appear to be high-level/strategic concerns. Noting the 'sensibility' and 'inherent' need to share is an important notion expressed by participants. Arguably, GISc is a discipline well-placed to expose and tackle the issues of sharing research data that JISC is exploring because of the raft of actors involved over several years of effort. This can, partly, be attributed to the need for nearly all GI to be created from the manipulation of data from different sources, the reliance on core data to enact and 'spatialise' datasets and the efforts and mission of the OGC.

Particular examples of this come through the notion of space as a means to represent data through its co-location as a basis to collaborative or inter/trans-disciplinary research. Many major social and environmental problems are better understood by this mixing of data in this way and, arguably, a lot of other practices would benefit from mapping their data, if only to improve data quality through visualising its geography[24]. The possibility of creating such large data assets for the use by many groups of researchers ties into a related area of e-(social) science.

As an emerging activity, this online infrastructure has collaborative research at the core of its definition, alongside the use of high power computing resources such as the Grid. Arguably, GISc should be the area of research that is building a large portion of this virtual or *in silico* research environment,

[23] see http://www.kazaa.com

[24] (such as instances where records appear to fall outside the geographical areas they were supposed to be gathered for. This is not an uncommon issue experienced by GIS users)

given the genuine demand for these resources and creation of large datasets that inherently (and already) require collaboration. If anything, there is a need for wider and deeper exploration as to why GIS-users would or should share their data, alongside what other information is required to foster this sharing. For example, one participant mentioned the pride she had in the datasets she created and that this would motivate her to share it, something seldom mentioned and, if encouraged culturally, that could counter some of the trust/control problems present in sharing GI.

5.2.5 Barriers to geospatial information-sharing

Barriers to sharing emerge in many more dimensions than drivers and relate, in part, to the critical nature of participants' views. From the outset, participants questioned others' willingness to install a technology to help them share datasets that exceed the (technical) limits of e-mail, for example. They felt there were difficulties relating to the naivety or ignorance of their community about 'sharing' GI and that practices were, in fact, "messy", limited, not common or occasional, there was low levels of adoption and, until the experiment, somewhat "untested". The lack of a critical mass of participants and datasets was also highlighted, with some questioning if too much of a cultural change would be needed to make sharing more acceptable or commonplace. Where sharing did occur it was felt to be time-constrained and "loose", perhaps impacting on the readiness to identify grey-sharing in practice, although a notion of sharing being "monitored" related this to issues of 'privacy'. In addition, the co-sharing of data and analyses between public sector agencies and academics has close ties to European policies relating to the creation of a knowledge economy through the reuse of public sector information and the importance of increasing access to information to aid decision-making. This topic is discussed below in terms of problems of the future evolution of the GRADE demonstrator repository.

Informal methods require a level of technical ability and additional effort to be a mechanism for sharing. Although participants felt this was something that would be easy to overcome, the ideas that the uptake of technologies may differ with certain groups (e.g. younger people adopting a series of technologies) was notable. Participants felt that additional effort would be needed to send information using the social network technologies, particularly if the recipient was not already signed-up, presenting a barrier that would lead to them to select alternative approaches. At an extreme, participants felt that limited Internet access was not common, and would only become a challenge in instances of mobile access in rural locations or if developments prevented their ready ability to find "fresh data" that was not being made accessible online.

The group saw data-policy barriers coming through license restrictions and 'rights'. They signalled digital rights and copyright as potential barriers but that problems could also be data-dependent. Participants highlighted the increasing interest in IPR, where they would not share data if rights were not clearly stated or if their university would not want them to 'give away'

valuable assets. This contention was also discussed in relation to the care needed in specifying the IP of datasets and how the negotiation of these rights is managed in relation to protection and acknowledgement/citation of dataset owners. Strategic research governance issues also emerge as barriers in this context, where the need to make participants and co-researchers aware of how data will be subsequently handled is emerging through greater restriction in sharing practices through research ethics and a desire to include sharing policies in research contracts (and their related infrastructures, such as the VKP resource).

The previously noted idea of citation and acknowledgement is better viewed in terms of issues of 'trust' and control. This topic is often recognised in information-sharing contexts but is somewhat difficult to readily characterise. Concerns were often expressed in terms of data accuracy and the nature of the potential audience for a given dataset. For example, participants were happy to share with most other GIS-users, but particularly those who were authenticated. They also raised issues of trust in terms of the data present in a shared resource and how to prevent users uploading incorrect data. This notion of a need for quality assurance also came against ideas of control, where a participant suggested varying levels of access to their dataset could come through different versions, with a shared version being somewhat incomplete and, therefore, potentially inaccurate. The fears expressed in relation to misuse noted a variety of concerns, where some raised concerns about being responsible for the data they provided and if they had any liability in relation to it. Data dependence also emerged from discussions, where a 'pure' OS dataset may be less problematic, in terms of liability, compared to the potential misinterpretation of derived data.

Solutions in this context also illustrated problems of trust, where participants wanted better control on the data through varying access levels, differing read/write permissions, digital watermarking and a system to report abuse. A more transparent system, around the *in silico* research approach was felt to increase transparency and have a positive impact on trust. Limited information of this type, and specifically lack of metadata, were felt to harm attempts to establish trust around the data and between users. A challenge remains in researching the extent to which trust is mediated through interaction/communication and collaboration and the potential transparency of online research, given the issues surrounding other e-developments, such as the 'digital façades' of government websites and the varying views of the actors involved in establishing and consuming them (see, for example, Smith, 2006). Certainly, informal repositories appear to have little role in mediating trust for wider consumption of data but more formal resources may offer greater opportunities.

Metadata was seen as key to current information-sharing practices. This element emerged several times and was tackled directly in terms of barriers, where metadata would be limited or missing. Participants felt that metadata is necessary to determine the quality of a dataset but that other information is

needed, again following the *in silico* notions of associating a dataset with its creator, purpose and applications/research outputs (such as publications), for example. A lack of metadata was felt to delay the sharing process, partly given the effort required in creating details from 'scratch'. The informal setting of sharing was also noted in this context, where less formal methods led to less formal descriptions of data. For example, in instances where variable headings would be missing from attribute tables, a definition from the provider (or the creation of their own "standards") in an e-mail or subsequent discussion would be enough to enable the recipient to make use of the data. Wider dissemination of the data would require additional information, partly linking to ideas of semantics around GI, something which is beyond the scope of this present discussion. Such barriers were also addressed through the SWOT analyses presented by the participants in the workshop, looking at the weaknesses and threats across both the informal social network methods and the pilot GRADE repository.

5.2.6 Problems with AllPeers, Exaroom and the GRADE demonstrator repository

Problems in this context can be considered from issues around the technology, its use, data and metadata, relationships between security/privacy and trust, the overall view of the GI community of sharing and the adoption of certain technologies for this purpose, alongside the impacts of alternatives.

The level of technical skill needed to use these technologies was emphasised in the SWOT. Participants outlined the challenges involved in the installation of the informal methods, partly in relation to changes in versions and the length of time taken to do this for those who did not have the DotNet for Exaroom or plug-ins for Firefox for AllPeers, as well as the time-consuming difficulties of adding friends. There were also concerns about relying on third-party software. These involved the sustainability of the products and wider issues of identifying the companies involved and what their software could potentially do to participants' systems. Participants also felt that the informal methods, in particular, involved a "cumbersome" approach requiring continuous login for the opportunity to access data/users. In addition, queries about the interoperability of both formal and informal methods were raised, with the GRADE demonstrator repository being able to interact with OGC web services, a potential weakness for future development.

The use and usability of technologies were also raised. Few participants considered the interfaces: some not liking the feel of AllPeers, others disliking that of the GRADE demonstrator repository. The confusion that some felt around the two informal methods on offer and the need for communication tools were notable, where the GRADE demonstrator repository lacks this opportunity to negotiate information about data (e.g. presence of a dataset or its online location) for some forms of (mostly) 'synchronous' sharing. Although participants noted that the technology was easy to use it may not necessarily be conducive to; where the pilot repository was felt to be time-consuming,

more formal and less attractive, leading to queries about the impact of under-use and the issues surrounding those who only engage in data extraction rather than input. As such, participants noted that the informal repository methods relied heavily on users to run and maintain them, something that the GRADE demonstrator repository may also face to some extent. This is further considered in terms of the project's current status, where participants highlighted the resource's sustainability relating to funding/resources and costs, the maintenance of these systems and the impact of the project's "shelf-life".

Data and metadata was also understood to be a problematic feature in the development of the sharing infrastructure in five areas. Firstly, variations in geospatial data-types were seen as weaknesses in both formal and informal cases. The current GRADE demonstrator repository is limited in terms of file types, with some non-database data not being 'acceptable' (such as images of maps or non-tabular GI) and that submission involved a slight technical hurdle involving the use of zip files. Secondly, the quantity of datasets was noted as a potential threat to GRADE as it could be a victim of its own success, impacting on the time taken to search through a larger catalogue of datasets and the tools needed to aid this (something later noted as an opportunity). This was tied to a related third area of the organisation of data that the informal repository methods lacked overall and which the formal repository lacked in terms of linking between datasets and the limited opportunity for users to create 'profiles' of available data through themes or topics. A fourth area of information management came through data discovery, which was seen as a challenge in real settings, and the ability to link through to resources from desktop GIS. Lastly, current activities around metadata were seen as a particular concern. For example, the need to repeatedly create content was seen as a specific weakness, alongside the inability to handle previously generated metadata content (such as information stored in ESRI's ArcCatalog). Threats were noted through the lack of sufficient metadata to judge the suitability of a dataset and the apparent limited use of metadata standards to help with the automated or coordinated retrieval of data.

Security, privacy and trust again emerged as a complex mix of issues from the participants. Threats were identified in the informal repository methods in terms of the invasive nature of the technology, breaches to firewalls or a general concern that central IT service departments are normally against peer-to-peer software present in the informal methods (partly due to the bandwidth they consume). This concern can also be seen in an extreme case through the perceived potential for "surveillance" through these systems or that participants queried the ability to trust system administrators or the real person behind the "Major Tom" avatar in Exaroom. Privacy emerged through concerns of general "IT paranoia" in the relatively computer-literate GISc community and that privacy was not only considered in terms of personal information or computer content but also around the presence of the data itself.

Specifically, the lack of control over data was an issue in both informal and formal methods with participants raising an interest in who was downloading their data. Trust issues also emerged around who was involved in sharing activities, from the specific examples in the experiment (where one participant's user-name was not recognised by others and, therefore, rejected as a 'friend'), to general concerns about knowing who a sharer is and attributing value to their datasets accordingly. This linked notions of trust to the quality of data directly, a lack of quality assurance, limited quality control (particularly for the wider infrastructure of the formal repository), which was subsequently tied to the idea of having the meaning of a dataset being expressed through community-mediated semantics. Again the lack of a facility for acknowledgement, citation and digital rights were noted as potential threats to sharing through the GRADE demonstrator repository in this context.

More general concerns relating to the informal repositories included querying the extent to which they were too focussed on social networks and sharing other information with friends, separating work practices and technologies from non-work ones. As such, it should be reiterated that such technologies currently are unlikely to be part of the grey-sharing of GI. Fundamentally, participants acknowledged other social forces in the selection of these technologies, noting that users have personal preferences of how they choose to share information and that the activity overall would be linked to the culture of sharing that may or may not emerge. This was queried further in terms of concerns about the growth of this community and that, indeed, many of the potential users of informal methods ("academics") and their current practices are perhaps the greatest barrier to information-sharing overall. This was emphasised in the context of the GRADE demonstrator repository in terms of its core purpose possibly not reflecting research practices, where data is highly valued at the time of production for the purposes of, for example, journal article publication and has diminishing value over time[25]. This also indicates notions of trust, where the lack of sharing can be related to limited trust of other researchers to use a dataset in competition with its original author(s) and, perhaps, the idea of introducing citation to acknowledge the source to mitigate this perceived threat.

Notably, challenges remain to any technology in terms of the potential alternatives. It has already been noted that current practice draws on readily identifiable and widely-adopted technologies such as USB storage, FTP, e-mail and burned media. It was also noted that threats to Exaroom and AllPeers came from more established social network tools and other web-based resources that the community could adopt (partly as they are maybe more known or 'tested'). In addition, the informal methods were contrasted with the GRADE demonstrator repository itself, with participants feeling it was needed, had a good, trusted background through EDINA and that it would

[25] until it increases in value as a historical artefact, as noted in a recent meeting involving digital curation. See presentation by Janée (2006): http://www.nesc.ac.uk/talks/697/ngda_Janee.ppt

foster sharing and trust through its centralised nature. In addition, participants felt that authentication through Digimap helped to handle some digital rights issues and that the formal approach was overall more stable, accessible and reliable. It should be noted, however, that a final threat could be present through the competition of alternative repositories, with NERC's facilities helping to maintain (solely) environmental datasets but that local resources or registries of dispersed online sources may also have a potential challenging or perhaps even contributory role.

5.2.7 Possibilities with AllPeers, Exaroom and the GRADE demonstrator repository

The technologies involved with this study were seen as benefiting from their overall low cost and potential adoption, particularly as a place where experiments could take place, their ease of use and accessibility, notions of community, interaction/communication and collaboration and links/contributions to wider infrastructures.

In contrast to the problems discussed above, participants seemed to favour both formal and informal methods in several ways. Both were valued because they were low cost, as the software did not have to be paid for but that some resources were needed in their use. Participants also saw merits in their technical features, including platform independence and being Internet-based technologies. As indicated above, very positive views were expressed about the potential of the GRADE demonstrator repository and participants noted that the intention for it to be a moderated GI-specific resource was welcomed, although possibly based on a technology that could not handle GI in interoperable or accessible way. Participants felt it was a experimental space for these issues, that it was a new activity, that nobody was had determined the best way to offer such a facility to date and that it was a timely opportunity to begin to develop best practice in sharing GI based on this resource. They noted the GRADE demonstrator repository's potential as a place where new facilities could be trialled and developed, highlighting the potential for a large number of datasets as an opportunity to develop approach GI search and retrieval tools. Comments also acknowledged that, to some degree, a lot of the technology is available today that could make sharing possible. This included suggesting that the informal methods offered enhanced versions of current practice and that peer-to-peer technology would be an advantage to the exchange of large datasets compared to e-mail, which also suffered from corrupting data. This again highlights the idea that sharing GI and the technology involved can relate to the purpose of sharing and that informal methods are currently less likely to be useful as a deposit and extraction approach *per se*.

Ease of registration for both informal and formal methods was notable, with the GRADE demonstrator repository's link to wider infrastructures (such as other EDINA resources) being a particular benefit for single user-authentication and ideas of more advanced technologies to aid this such as Shibboleth. Participants also acknowledged that, in general, both formal and

informal methods were easy to use and that the GRADE demonstrator repository was currently straight-forward to add and search data contained at a single location. Exaroom was highlighted as having a good user-interface and appearance, partly because it involved several other features beyond other than just the means to share data. To some extent this included more social aspects, such as the ability to upload images/photographs to avatars, something that some participants chose to do. It may, therefore be interesting to explore other social issues relating to sharing such as the extent to which presenting an online persona impacts on notions of trust by literally 'putting a face to a name' when online.

Accessibility was also considered as a positive characteristic of both sets of methods, where speed and immediacy were possible for some sharing using AllPeers (in particular) and that the GRADE demonstrator repository was readily accessible and "always on". The control notion was also noted, again, in this context in terms of being able to restrict who were 'friends' in the informal methods (although some concern was raised about the potential to browse the folders of friends-of-friends). Such a desire for restriction to the data introduces a number of management issues, although licensed/registered users were noted as being of less concern in the first instance. How controls could, then, be put in place to monitor an authenticated group's use of others' data was also a concern of the participants. The clear problem for this form of control is that once a dataset is outside of the repository it can be passed to others diminishing/negating all security. Again, an *in silico* approach would suggest a need for resources that allowed the processing of data within the secure environment without any need for data to be removed from the resource. It is suggested that this is a challenge for both technical and socio-cultural issues surrounding current research practice, let alone for the technical and financial resources needed to develop such a facility.

Continuing on the e-(social) science theme, participants valued the notions of community, interaction and collaboration highly in the SWOT. Community-related aspects included the sense of a virtualised community and notions of user-visibility that the informal repository methods could foster, as well as information-sharing communities that would help to expand personal networks of GI users. Such a development was also felt to help involve more potential users, to an activity dependent on the contributions of many data assets that may not be recognised as GI by their current owners. In comparison, interactive facilities were noted in terms of communication, in particular, with active forms including the use of IM in AllPeers to help negotiate the technical aspects of sharing (and help manage the experiment in practice); a sense of immediacy in communication through these tools; and the desire for more tools to enable simultaneous interaction amongst multiple users. More passive forms of interaction appeared in terms of browsing remote machines or the idea of producing a resource as "frequently asked datasets", particularly in teaching contexts. This could involve assets such as core/framework data and different types of interaction, varying from

asynchronous to (almost) synchronous sharing (with or without prompting) and recognising the degree to which the sharer needs to be active or passive, although the handing-on of data in teaching contexts may provide one example of grey-sharing that does take place[26].

Ideas of interaction also came through the potential for the informal methods to be used in group projects alongside an idea that such resources would help projects to be resilient to staff change if there was a common place to maintain data. Another opportunity existed in this interactive context in terms of the creation of shared data by augmenting official sources with new data, correcting any errors in these sources and creating new GI from public sector information, in particular, as well as providing a place for orphan datasets to be tested and, possibly, 'fostered' through re-use. Such ideas of interaction also bring in notions of increased use of data leading to improved quality.

AllPeers, Exaroom and the GRADE demonstrator repository were labelled as "collaborative" or "collaborative methods" by the participants. Opportunities were felt to exist to aid quick negotiation in a collaborative setting, although the GRADE demonstrator repository currently lacks interactive tools. However, it was valued as a place where inter- and intra-institutional collaboration could take place around data, with the potential to produce a shared resource that could reduce or remove duplications of effort. In particular, participants identified the possibility of creating collaborative teaching and research repositories, with the latter seen as a potential resource that could be used by non-academic collaborators (such as public sector information providers and researchers). However, as noted above, the lack of this mutual resource for the wider GI community was felt to be a threat, something beyond the main scope of the GRADE project.

As well as the aforementioned potential as a major data asset for e-(social) science and a component of an expanded infrastructure, the GRADE demonstrator repository was highlighted for its links to two other resources. The first is the possibilities offered by connecting the data to Go-Geo!, where metadata would be made available about assets in the repository and where more communicative/community tools could exist around this metadata. Secondly, queries relating to wider collaboration link to what the development of the UK National Spatial Data Infrastructure, least of all in terms of ensuring definitive datasets are made accessible to researchers collaborating from different sectors, removing duplication of effort and making potential GI accessible and, perhaps more importantly, used.

5.3 Conclusions and Recommendations

To summarise some of these points, the ideas expressed by the participants in relation to workshop tasks help illustrate some more opportunities to both

[26] (possibly as this is authorised through OS licensing, although this still presents issues for the uptake and consumption of any data assets provided by EDINA, for example)

the GRADE demonstrator repository and geospatial information-sharing, in general. Participants were very much in favour of having a national geospatial data repository, possibly linked to wider resources such as SDIs. They were also keen for a central portal, possibly linked to Go-Geo!. The creation (or maintenance) of local or departmental resources were valued in different ways, with some participants feeling they should move from this position towards a national repository, that they only had a role as back-up or that there may be some value in a mixture of local-to-global information-sharing, given the principles outlined in recent GI policies such as the new INSPIRE Directive[27]. In part, this area also raises questions about the extent to which academic information should play a role in this pan-European infrastructure and whether the academic community are being viewed as 'citizens' under INSPIRE's current information-sharing model. Participants did not value lists of informal data sources as highly, partly due to them appearing as less robust or reliable on their own. However, it was felt that such informal repositories could facilitate access to more information by allowing extracted and 'cleaned' data to be moved to a repository and for registry web services to make such distributed data more accessible (again potentially following the local-to-global model) and building on a related desire for the implementation of OGC standards in this context.

In terms of licensing, participants felt that a reduction in restrictions had some advantages but either felt it would not happen or queried how this could be done, whereas removing licensing restrictions was seen as not appropriate. Licenses were valued as they were felt to help protect both users and producers of data. Experiences from the experiment raised the invasive nature of a GIS data forum like Napster and concerns about validating who would have access to the users' datasets. A free market on data does not appear to be desired by the participants in this case at least. This should also be considered in terms of the participants' varying familiarity with such *fora* but also the extent to which having a large number of users involved may impact on people's willingness to share.

A couple of more positive positions also emerged, where such systems would offer a chance to view more useful datasets and provide an opportunity to explore more *in silico* research activity around the labelling and discussion of datasets, tied to social tools such as rating and recommender systems. Such ideas reflect on participants' interests in who uses the data and what they are using it for. Some felt that they would only want such a facility out of curiosity but others related this to being able to control the use of their data, alongside research governance issues of being clear with research partners and participants in how data could be maintained and re-used. Lastly, most participants felt that increased computer speeds would be inevitable (for the time being) and that such issues only impact on large datasets, something which GI research frequently creates and an area that e-(social) science is being developed to manage.

[27] The new European geospatial data-sharing Directive, see http://www.ec-gis.org/inspire

By lacking a theoretical approach to this material, it is difficult to draw firm conclusions about the notions being expressed. What this has allowed, at least, is the exploration of the topic across several themes that some 'representatives' from the UK GI community have identified themselves, possibly lending this material to a grounded theory approach, where theoretical ideas emerge from the topics discussed. However, it is suggested that it would be better to view this report as a means to highlight topics that need further 'unpacking' and 'testing' in relation to both social theory and technological development.

For example, 'trust' has been identified as a concern closely associated with shared data, alongside the subsequent re-use of data by other GI Scientists and a desire to control its use. Questions remain about the extent to which communication fosters trust or mitigates risks around data re-use through metadata or the *in silico* research environments/social tools of e-(social) science. Do such online resources present another 'digital façade' in the wider information society? Are they a necessary 'stepping-stone' to building either ever tighter (and more monitored restrictions on) data re-use or a much more open and community-driven resource? Certainly, the strategic view that many participants took in viewing this topic is notable. The experiences gained in participating in the study have let to a form of action research to take place. This needs to be built upon and expanded to include other stakeholders, as varying local circumstances and personal preferences have been highlighted as potential impacts on the selection of methods to share GI in the academic sector.

Given these concerns and possibilities, some recommendations can be made based on the activities from this study:

1. There is a need to address the GISc community's concerns and possible mis-conceptions about licensing restrictions against a need to share data. Certainly, the work carried out in WP3 of GRADE relating to digital rights and licensed users should help in what will need to be an educational and participatory process[28].
2. Currently, GIS-users appear to have a mixture of sufficient approaches to share data, in general. Informal repositories could have some role in geospatial data-sharing for small group activities but they appear to have limited utility to act as a distributed national resource (except as in recommendation 7). More work is needed to explore and monitor the uptake and role of informal repositories in small group settings and how they could contribute to a wider infrastructure.
3. The development of a national geospatial data repository was well-supported by the study's participants and should be promoted. As such, there is a need to identify champions in local settings to promote and encourage its use. Regional training sessions and promotional

[28] see http://edina.ac.uk/projects/grade/gradeDigitalRightsIssues.pdf

resources should be developed to fit varying audiences that are likely to emerge in both research and teaching.

4. There is a need to develop links between a national geospatial data repository and other EDINA resources. In particular, consideration could be given to the core/framework datasets held in, for example, Digimap and UK Borders and storage issues that may otherwise emerge through duplicated geometry. Such an approach would also aid comparisons of the spatial extent of topics, both in terms of the patchwork of geographical coverage of a given theme or the coverage of themes at a given location.
5. Similarly, the establishment of the GRADE repository should be considered around the role of Go-Geo!, where EDINA should envisage Go-Geo! as a geospatial 'one-stop-shop' for the UK academia. In particular, the service should help users to find, evaluate and re-use geospatial assets held within a national facility. Go-Geo! should also be capable of searching and accessing geospatial data in any repository capable of managing licensed geospatial data as well as exploring the provision of the social/community tools seen to be needed to negotiate, discuss and, potentially, instil 'trust' around GI, towards *in silico* research.
6. As such, there is a need to consider the metadata that this should involve, drawing on information that users have already provided in existing software. In addition, there is a need to raise awareness of formal metadata for both sharing and personal data- maintenance/-curation, something requiring greater promotion at training and grassroots levels, beyond just the GISc community.
7. The creation of an accessible permanent central geospatial data repository offers opportunities to link to wider data sources. The relationships between a national central repository and formal institutional repositories, alongside grey-sharing and the fostering of orphan datasets, should be explored in terms of OGC compliant registries and the means to draw neglected data into a place where it can be stored, re-used, validated and valued.
8. The creation of a successful national geospatial data repository offers opportunities to develop Geographical Information Retrieval tools capable of searching and making sense of both a plethora of data and, through middleware, within massive datasets. Such ideas also link to the challenges of currently non-standard forms of GI being considered as part of the semantic web and how (community-mediated) semantic and technical interoperability play a role in having users readily access the data they need.
9. Consideration should be given to current research practice and the desire to develop a wider infrastructure in the context of *in silico* research and NSDIs, particularly through the conditions that would need to be put in place to allow authorised access by public sector colleagues to metadata and dataset 'discussion' and, if appropriate, allow authenticated access to the actual datasets, especially given the potential roles of academia in activities such as INSPIRE.

10. Such activity should be recognised as an opportunity to continue research and exploration involving longitudinal and *in situ* qualitative approaches to better understand the demands of a variety of GIS-users, in relation to concerns about trust relating to their data and that of others, and the development of appropriate theoretical models to understand the social and technical context and role that (academic) geospatial data repositories play in Europe's wider 'e-society'.

References

Bryman, A. (2004) *Social Research Methods*. Oxford; Oxford University Press.

Commission on Geosciences, Environment and Resources (CGER) (1999) *Distributed Geolibraries: Spatial Information Resources, Summary of a Workshop*. http://books.nap.edu/openbook.php?record_id=9460&page=R1 11/04/07 1726h

Dingwall, R. (2006) Confronting the Ant-Democrats: the unethical nature of ethical regulation in social science. *Medical Sociology Online* (1) pp51-58. http://www.medicalsociologyonline.org/current/pdf/conf_anti.pdf 11/04/07 1530h

McKee, L. and Pichler, G. (2003) Registries and e-Services: final report. GINIE Project Deliverable 2.6.2 (b) 14/08/2003. see http://www.ec-gis.org/ginie/doc/e-services_final_report.pdf 27/02/07 2300h

Smith, R.S. (2006) Theories of Digital Participation (Ch3) in Campagna, M. (ed) *GIS for Sustainable Development*. London; Taylor and Francis.

Woolgar, S. (2003) Social Shaping Perspectives on e-Science and e-Social Science: the case for research support. A consultative study for the Economic and Social Research Council (ESRC). http://www.ncess.ac.uk/docs/social_shaping_perspectives.pdf 11/04/07 1100h

Appendix 1: Informal Repositories Evaluation

As well as searching for examples online, part of the exploration of possible informal repositories for the study involved an online article assessing four examples (AllPeers, Exaroom, Pando and Zapr), with a running debate by readers and continuous updates of the development of peer-to-peer sharing technologies[29]. This appendix gives a brief overview of the technologies discussed in this material, leading to the selection of AllPeers and Exaroom for the experiment.

The review covers the following softwares and/or online services:

- Avvenu
- Box
- Dropsend
- ESNIPS
- FastSend
- Foldershare
- Geotorrent
- Google Talk
- Lionshare
- Neobebox
- Pando
- Pubet
- rarhost
- Tamago
- Windows Live Messenger
- yousendit
- Youswap
- Zapr

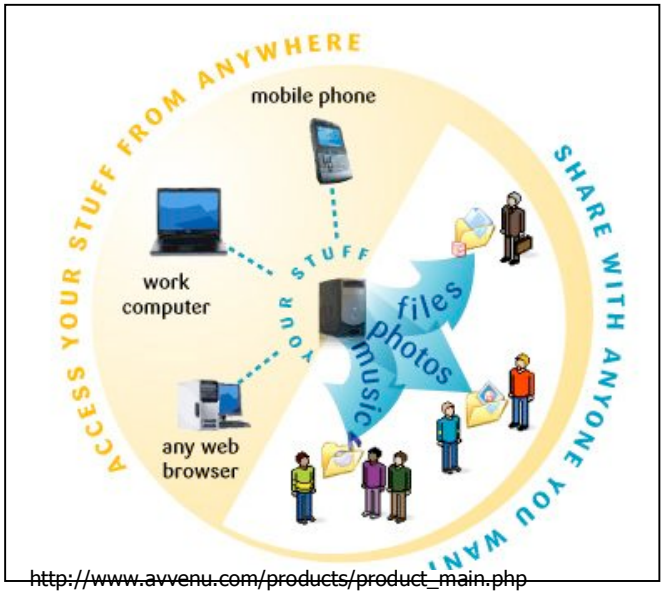
The discussion focuses on the following characteristics:

- If they are a beta software
- If they are free
- If they are a client or online resource
- Any file size restrictions (if known)
- Features for collaboration
- Problems impacting on their selection for this study

[29] see http://www.techcrunch.com/tag/AllPeers with reference to Arrington, M. (2006) *Let's Share Some Files - Four Services Compared* http://www.techcrunch.com/2006/08/24/lets-share-some-files-four-services-compared

Avvenu

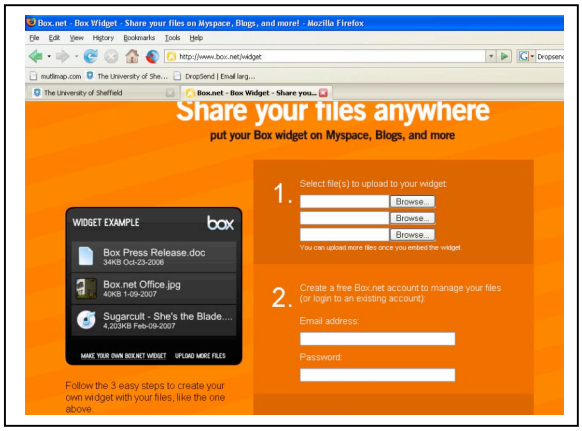
URL	http://www.avvenu.com
Beta?	No
Cost	10Gb space at $30US/yr or $4US/mo
Type	Online resource but with mobile capability too
Collaboration	Platform independent access to files for access and sharing
File size limits	unknown
Selection issues	costs for access when user's machine is offline



http://www.avvenu.com/products/product_main.php

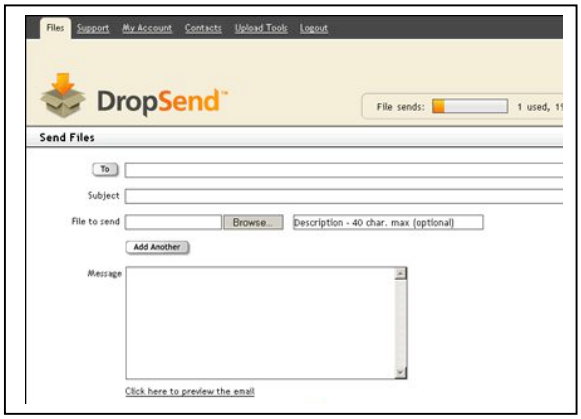
Box

URL	http://www.box.net
Beta?	No
Cost	5Gb space for $7.95 per month
Type	Online resource
Collaboration	Can make files or folder public for sharing
File size limits	Unknown
Selection issues	Although a free trial exists there are costs



Dropsend

URL	http://www.dropsend.com/
Beta?	No
Cost	Varied charges
Type	Online resource
Collaboration	Send or store files online
File size limits	< 1Gb
Selection issues	limited free version

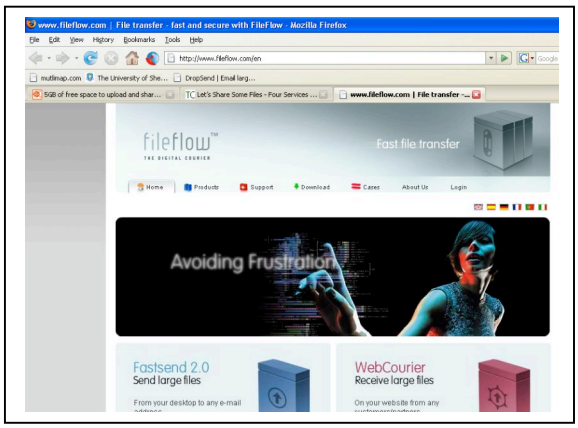


ESNIPS

URL	http://www.esnips.com
Beta?	Yes
Cost	Free
Type	Online resource
Collaboration	Online media sharing site (like Tamago)
File size limits	<5Gb (storage)
Selection issues	Aimed at selling media- no section just to share data

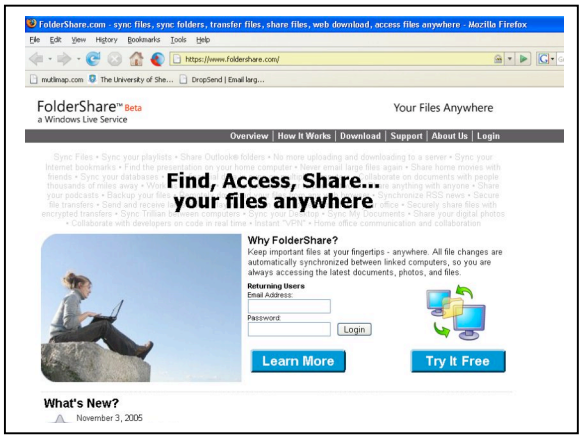
FastSend (fileflow)

URL	http://www.fileflow.com
Beta?	No
Cost	$240/yr
Type	Online resource
Collaboration	Can send and upload files linked through a URL
File size limits	Unknown
Selection issues	Although a free trial exists there are costs



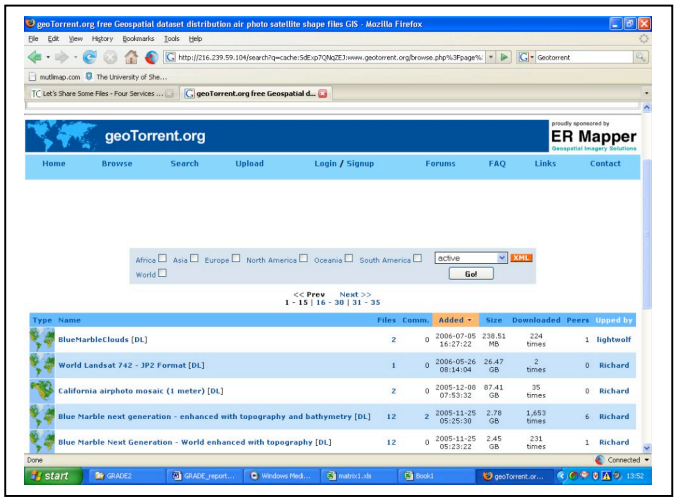
Foldershare

URL	http://www.foldershare.com
Beta?	Yes
Cost	Free
Type	Download data 'satellite'
Collaboration	Now part of Windows Live
File size limits	<2 Gb
Selection issues	File size restrictions may be an issue for GI



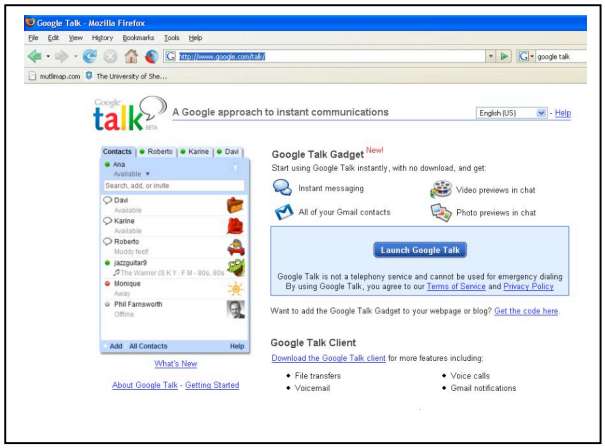
Geotorrent

URL	http://www.geotorrent.org
Beta?	No
Cost	Free
Type	Online resource
Collaboration	Forums and a search and retrieval section- GI specific
File size limits	> 87Gb
Selection issues	Could not register and server was down



Google Talk

URL	http://www.google.com/talk
Beta?	yes
Cost	Free
Type	IM with sharing capability
Collaboration	IM client that can share files
File size limits	Unknown
Selection issues	Limited to GMAIL users only



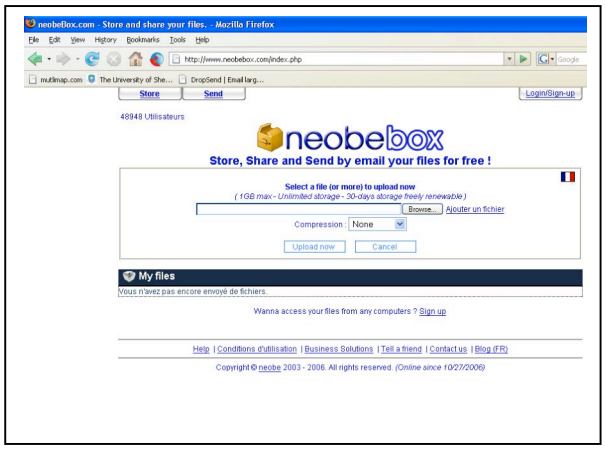
Lionshare

URL	http://lionshare.its.psu.edu
Beta?	No
Cost	Free
Type	Online resource
Collaboration	Similar to Kazaa and Gnutella but with user authentication. Designed for the academic community
File size limits	Unknown
Selection issues	Could not install because of security problems

http://lionshare.its.psu.edu/features

Neobebox

URL	http://www.neobebox.com
Beta?	No
Cost	Free
Type	Online resource
Collaboration	Upload straight from website files to share using permanent URL
File size limits	Unknown
Selection issues	Most of the instructions are in French, uploading is not clear if it completes. Like hosting from personal webpage



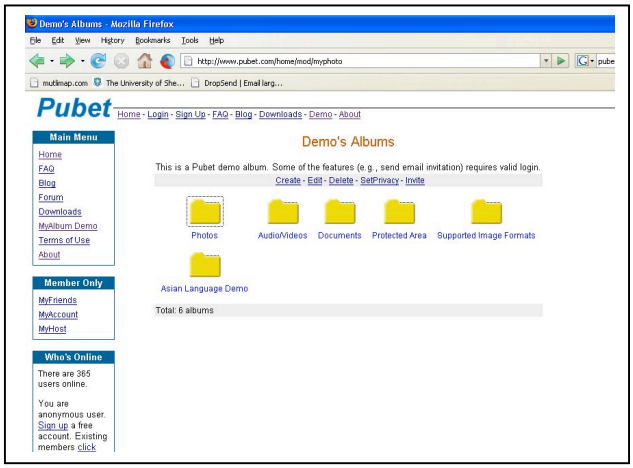
Pando

URL	http://www.pando.com
Beta?	No
Cost	Free
Type	Desktop application
Collaboration	Multiple recipients (all recipients also become senders)
File size limits	<1Gb
Selection issues	Managed to share some info but receiver could not re-send

http://www.pando.com/what

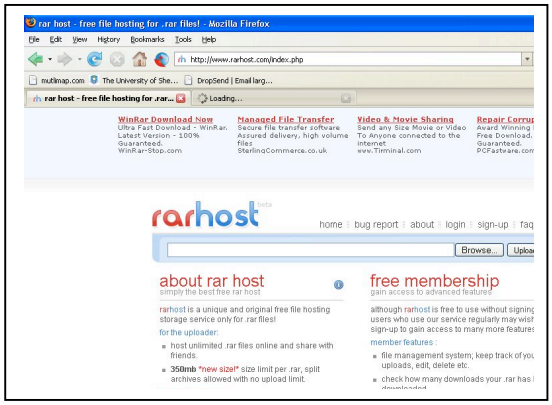
Pubet

URL	http://www.pubet.com
Beta?	No
Cost	Free
Type	Online resource
Collaboration	Different sharing modes: both for public/private sharing. Similar to Exaroom
File size limits	Unknown
Selection issues	Interface seems dated



RARhost

URL	http://www.rarhost.com
Beta?	No
Cost	Free
Type	Online resource
Collaboration	File hosting site for RAR files (only)
File size limits	<350Mb
Selection issues	Test did not seem to complete. Files only available for a short time (1-2 months). RAR compression only



Tamago

URL	http://www.tamago.us
Beta?	No
Cost	Free
Type	Online resource
Collaboration	music/arts/art community, has digital fingerprinting
File size limits	Unknown
Selection issues	'Commercial'? Too many additional 'social issues' for this study- may be useful for another

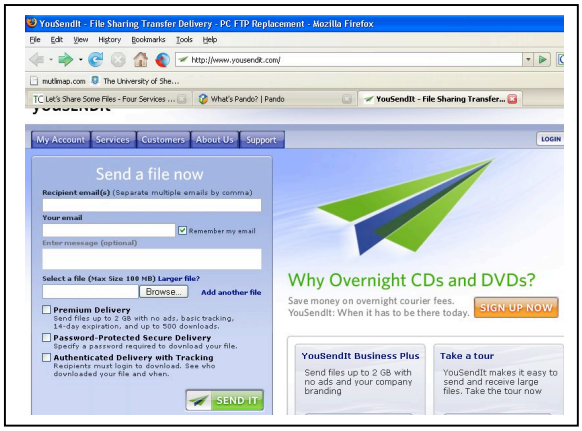
http://www.tamago.us/download/index.htm

Windows Live Messenger

URL	http://www.msn.co.uk/livemessenger
Beta?	No
Cost	Free
Type	IM with sharing capability
Collaboration	IM client that can share files
File size limits	Unknown
Selection issues	Widely used for communication but 'additional' to data-sharing?

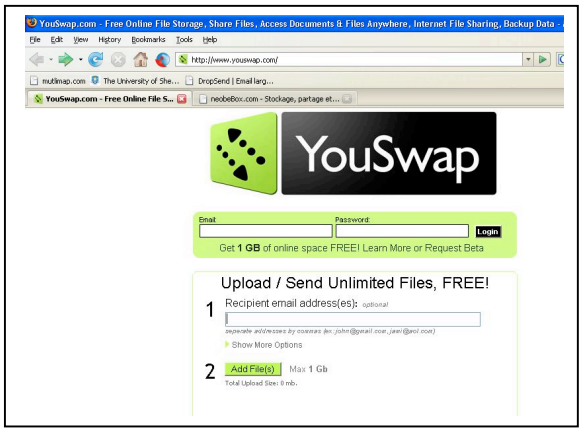
Yousendit

URL	http://www.yousendit.com
Beta?	No
Cost	Free
Type	Online resource
Collaboration	Upload and limited number of downloads (25x in 7 days) for the free version
File size limits	<100Mb for free, can pay upto $29 per month for larger files
Selection issues	Arrington (2006) noted limited functionality



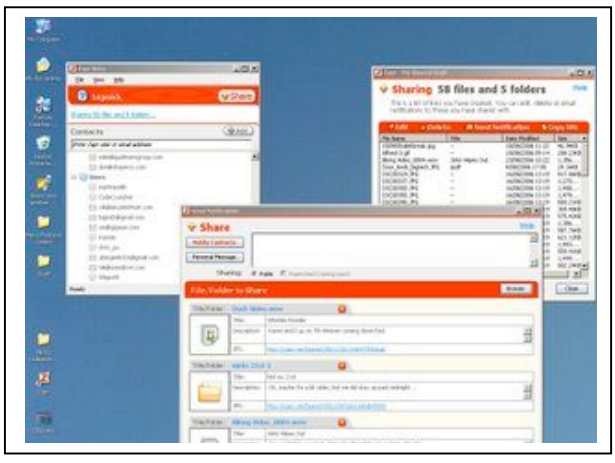
YouSwap

URL	http://www.YouSwap.com
Beta?	Yes
Cost	Free
Type	Online resource
Collaboration	Upload straight from website files to share to e-mail addresses
File size limits	<1Gb
Selection issues	lack of communication but very easy to upload (also monitors number of downloads)- '*reserve' technology for experiments*



Zapr

URL	http://www.zapr.com
Beta?	Yes
Cost	Free
Type	Downloadable application
Collaboration	Drag files into Zapr, send to email address or Zapr username
File size limits	No limit
Selection issues	Mainly 1:1 transfer. Slows computer and could not share data



http://zapr.typepad.com/media