

## General comments on data management

Please feel free to add any other comments regarding data management, long term data storage and access, digitisation, training, etc.

[-----] My most pressing need is greater storage capacity and the ability to share data with researchers abroad. The [university] e-mail system is entirely unsuitable for me to send data to co-authors abroad. In addition, the limited storage capacity I have means I have to manually back up my data every so often to an external hard drive. Sometimes I forget especially when I am busy before a trip which is not good.

[-----] The data archive is so large, it is physically difficult and expensive to manage.

[-----] There is severe lack of space on our network for storage. It's like being in the dark ages.

[-----] The default storage (100 MBytes) for emails through [university service] is absurdly small considering I can be sent in the order of 20 to 30 Mbytes a week.

[-----] I would like to have a version management fro my papers - we used to have in Germany a server run by the uni where I could upload my files and got them from there every time and I could access with login from different location without the need to create a VPN.

[-----] I am not a 'major' data player.

[-----] My data is qualitative and while I would be happy for it to be shared, there are some ethics hurdles that would need to be overcome

[-----] We need a central site what gathers and sorts a range of data on a geographical scale. the number of times people must download the same data from ABS, etc and then construct their own datasets is too numerous to mention.

[-----] With respect to the above question, I'm not paid enough to get involved in this (half time, level B, soft money)

[-----] our support facilities for long term data storage and management are minimal.

[-----] The University should be aware of the "Data for Science" work of PMSEIC and support it the link is as follows

[http://www.dest.gov.au/sectors/science\\_innovation/science\\_agencies\\_committees/prime\\_ministers\\_science\\_engineering\\_innovation\\_council/meetings/sixteenth\\_meeting.htm](http://www.dest.gov.au/sectors/science_innovation/science_agencies_committees/prime_ministers_science_engineering_innovation_council/meetings/sixteenth_meeting.htm)

[-----] Currently (at least in my school, [...]) there is a lack of a good "archival" data system. Need access to storage other than the frequent use main server (which is full anyway) so that we can put data into a permanent (infrequent access) repository server or tape based system. Having our data on dvds alone is a bit scary - easy to drop and break, or get lost when staff change etc!... we need better systems for labs like mine that generate lots of high res images etc.

[-----] Apart from keeping back-ups of all work, feels like mostly applicable to people who work with primary data.

[-----] We have very limited access to secure network back-up space in nursing which I consider preferable and more secure than CDs etc

[-----] How will this particular project integrate with existing national databases, and with personal lab webpages?

[-----] As long as it was not a time consuming process. I probably haven't got the volume of data to be useful to a reference committee

[-----] I require that my PhD students have filing cabinets that lock for the long term storage of raw data. Some people feel that provision of lockable cabinets is not a research expense that should be subsidised by the university, even though the students have a work desk on campus.

[-----] [The university] is a little behind on this front, as compared to other organisations I've worked for.

[-----] The University can create a website data base and allocate space for all the schools. Every school should allocate space for each group. Every group should have allocated space open for everybody (official website) and space allocated only for people with special permission. In this manner everybody will have access to his data from everywhere (non-official website) and will have contribution with his official results to the official website of his group. Every group should have a manager of its website who will put in order the information and make it attractive to external visitors. Advertisements of the group can be made and investors can be attracted in this manner.

[-----] Long term management of data can be left up to the discretion of the individual group leader of a lab. Short term everything has to be backed up or archived.

[-----] Erk, is the best I can say. Sorry to seem hostile, but your questions don't really bear much

## General comments on data management

relation to my work though I hope they are helpful to other people.

[-----] It is too hard for individual researchers to run large digital data sets. We already carry large workloads and there is no use having data sets when you get no time to publish. What might work better is to look beyond fixed data sets and link data collection in the social sciences and humanities to the work the library does. A researcher could go to a librarian and specify a data set (e.g. video, pictures, electronic texts, statistical data, etc.) that the librarian would collect for the researcher. It would also be useful to get easy access to people who can create the online interfaces (e.g. online surveys) that would draw data into the university. In short, the library could see itself as a mediator between data and researchers in the social sciences and humanities. Finally, there is a lack of understanding of the potential of digital storage and access technology on the social science and humanities side of the university and that makes getting involved in such projects too hard. Much of the work I have done (and now am keen to let go of) has been regarded as irrelevant and has been met with apathy (and often worse) by the senior colleagues I report to. My chief collaborator in [this part of the university] has found it more amenable to go and work [for another university]. Seen from an social science and humanities perspective, it is clear that [another university] is years ahead of [this university] in its approach to these issues. The largest data set I have involvement in is the [department's] data set and will shortly be closed down. It has more than 1 TB but once it is gone I will mostly only have textual data stored electronically.

[-----] I feel that it is very important to start training students in issues relating to data management including ethical issues along with technological.

[-----] I usually only gather a small amount of data, so most of my projects aren't relevant to these concerns.

[-----] Re Q15: I would be willing to participate subject to time limitations. We are situated off campus at [...].

[-----] [One part of this university] has LIEF funding to develop a prototype data archive to house digitised textual data. This will be a new node in the Australian Social Science Data Archives. The major challenge for the Australian Social Science Data Archives (all nodes) is how to secure recurrent funding. At the moment ASSDA funding has been primarily through local support and LIEF, but LIEF is not an appropriate mechanism for ongoing support for the archive.

[-----] Perhaps it would be useful for any eResearch reference group at [this university] to collaborate with teams of [this university's] researchers who are setting up eResearch facilities. For instance, we are currently getting ready to survey identified qualitative researchers.

[-----] As a Research only staff member, I don't feel that I have the time (or perhaps support) that I need to manage my data as it should

[-----] I am an organised kind of person, so I think I have data management under control for my own work. I stress to my students and postdocs the importance of keeping copies, and so far there haven't been any disasters.

[-----] would be good for you to talk with my research staff as well - they may have different needs

[-----] Data management bureaucracy seems unnecessary for social science research.

[-----] this is an area that staff in our School need assistance with - it is not given sufficient consideration at present by most staff. We lack training in this area

[-----] Our requirements will include longer-term data storage/backup as well as IT support for data management and access. These have to be integrated with specialized software to ensure functionality of our entire, multi-instrument genomics platform. That is our needs go beyond data storage and management.

[-----] Data must be openly accessible, anything else is a waste of time.

[-----] I don't believe our collected data will ever be available in the public domain.

[-----] University guidelines on these topics would be useful

[-----] this is a great initiative

[-----] Qualitative research management is important too

[-----] Re [next question] I don't know if I have anything to offer but would be interested in knowing more about the expectations of participation in an eResearch reference group

[-----] data management is a debacle within the university. if RQF outputs cannot be managed properly in [the system] I wouldn't be holding my breath that research data could be stored successfully

[-----] I have no idea what eresearch is

## General comments on data management

[-----] I believe more education, and ideally coupled with easily accessible backup/storage facilities, is required for long term storage for large datasets. Colleges of mine had an annoying incident of data loss several years ago, due to storage on recordable DVDs, which are not an adequate medium for long term archival purposes. This was in part due to ignorance of good procedure and in part due to lack of funds or facilities to maintain proper backups of large (>100 Gb) datasets.

[-----] As film and television production moves to file based acquisition (instead of tape and film) the data quantities to be managed are significant ( 1 second of uncompressed HD video is approx 178Mbytes) - backup, archive and management of petabytes of data is a task the industry is attempting to grapple with as we write.

[-----] I feel I have had no training in how to set up or manage data and am concerned about longevity of the data storage and platforms that I am using. I anticipate that I and others will still be using the data I am now collecting well into the future (as it is historical data, so does not date as much as other disciplines) and am unsure how to make sure it is still accessible.

[-----] I have worked in a multinational pharmaceutical company in the USA where we had a greater choice of ways to keep our lab books. The method which I was particularly keen on (and was very popular within the company) was to have access to individually numbered A4 sheets on which we were then able to print our electronically kept lab books for signing. Since the climate is such that signed lab books are very important for IP, I think that the university should seriously think about implementing a similar method. I think that it is recognised that it is quicker to write up electronic notes, rather than laboriously hand write lab books. As researchers we are under enormous pressure to maintain a high output. This would be an important strategy to allow researchers to maintain a high output as well as well kept lab notebooks.

[-----] The [institute] manages this data collection under a contract with the Australian Government. That contract requires us to manage all aspects of data management and security. We have a series of detailed plans for storage of the data (both in its original paper form and its various electronic formats), protecting the data, and retrieval of the data in case of emergencies. We also have joint responsibilities (with the Commonwealth) for distribution of the data and monitoring use of the data by the user community. There are over 800 registered users of our at least one of our 5 data releases.

[-----] Normally I would be willing to participate in a study, but I have serious health problems and am only able to work part time on light duties, and prefer to use this time to meet particular objectives.

[-----] [The university's system] is very poor for recording research outputs

[-----] I am not sure what eResearch is

[-----] I dont know what eResearch means - so the answer in [this question] is meaningless.

[-----] huge issue and not one i have adequately sorted in the lab environment despite years of trying - biggest issue is getting buy-in to manage data sensibly from co-workers.

[-----] Even though I've answered yes to the question above I'm not really interested. eResearch is just a tool we use. It is not the focus of our research. I supervise a PhD student whose project is heavily based in eResearch and I think she would be very intersted in being part of your eResearch reference group.

[-----] What on earth does "eResearch" mean?

[-----] The pressures of research, teaching and administrative duties mean that there is too little time to implement and comply with any extra burden that might be generated by formal systems for management of electronic (or non-electronic) research data.

[-----] The backup options provided via the Departmental servers are woefully inadequate - less than 200MB per staff member/postgrad is allowed, and there wouldn't be enough space on the disk if everyone used their quota. No training in research data management is provided, as the IT staff are too busy fixing breakdowns and installing software to pay any attention to advising people on how to back up their data. Members of our research group have bought their own (personal) USB drives to back up data onto.

[-----] All of these should be subject to an enterprise-level information architecture, supported by enterprise-grade IT architecture

[-----] too busy at present

[-----] As well as own research, I am responsible for the student research in our program. Want to create a system to store and manage all electronic files associated with the research (mostly SPSS, n-vivo, word and audio files) as a resource for secondary research.

## General comments on data management

[-----] Time consuming to insist that everyone does it, and to build efficient systems for handling data, but we must improve it for long term storage and data verification, security of info etc.

[-----] I am privileged to be given access to many [data sources] at a number of General Practices, where I set up clinical/information systems for Chronic Disease management, Data cleansing and coding, electronic clinical (EB) guideline templates, etc.

[-----] A significant problem in the work I do is recruiting informants, and an important concern to be met is that the purposes of the research be clear, the methods be clear, and boundaries set accordingly. Should the data I collect be made available to other unknown researchers, for other unknown purposes, using other unknown methods, I expect that potential informants will be justifiably less likely to participate.

[-----] I feel that the current informal data storage system works well for me.

[-----] I don't actually know what is meant by eResearch. My work uses digital data storage for recordings, images, video files, derived data. We also undertake computer simulations using realistic models and this generates very large data sets. This would not be possible without computer based digital storage, but the key is biology, the computer is merely the tool for organising data.

[-----] I would like to be able to make all my printed articles available for other researchers to use but not abuse but this is not currently on offer in my Faculty. As an older honorary working in Australian music history my articles will endure but need to be easily accessed by the public and all from the same site.

[-----] If any prescribed practices are set in place, please make them simple and painless (cf. my answer to Q 15).

[-----] this seems totally useless

[-----] I am not sure what you mean by eResearch.

[-----] I am responsible for the data I generate as part of the project that I am contracted to carry out. The issue of more global data management is better addressed at a higher academic level.

[-----] Clearly piles of CD's etc is an unsustainable strategy, but I am concerned about the tail wagging the dog here. Most journals encourage submission of datasets to their on-line repositories so the problem is beginning to go away (at least a bit).

[-----] I think that my comments above summarise my feelings

[-----] I think the issues regarding the storage of research data varies significantly between disciplines. The University seems to have a one-size-fits-all policy which is rather silly when applied to some fields.

[-----] I can appreciate the logistical and practical difficulties of backing up data--one is constantly trying to manage everything on the computer alone, never mind managing what was backed up but no longer needs to be, which backed-up files need to be preserved, and which need to be updated. Hopefully those thinking about this problem more than I have some ideas, but I would like to caution against singular solutions which are designed to suit specific interests but are then forced on everybody (a la [the local system], which I can imagine is an accountant's or auditors dream, but is a shocking hindrance to far more people -- "self-service" indeed). I would argue that the primary problem at the local level in the short term is simply space and time/convenience, and in the long term is data format. The overhead required to partition all my data and work into bits to fit on CDs or DVDs means I'll rarely do it, and these formats aren't for forever anyway. Also, external hard drives in the sizes we would need to be practical are prohibitively expensive. And I have spent ages in the past combing through 5 1/4 floppies to transfer to 3" floppies, only to spend ages again transferring to IOMEGA and Bournulli (sp?) disks, then Zip disks, then CDs and DVDs, etc. I certainly don't have the solution, but having a keen understanding of the problem makes me cautious of suggestions that there is an obvious solution.

[-----] I wouldn't be practising eResearch at the moment but as a research centre we have for about 2 years now explored what is possible re developing a database for the large amounts of qualitative data we have and that is generated by partners. We see this as necessary and important for a number of reasons. We have a small grant to explore the ethical and management issues of this at the moment but in the past we have been frustrated overall about the lack of IT expertise and guidance we have received in investigating this.

[-----] chemoffice offers enotebook. i havent implemented it, because of lack of training/time, but believe a web-based chemoffice environment would provide enormous improvement in communication with collaborators, nationally and internationally

[-----] interested in knowing about IP and long-term data storage

## General comments on data management

- [-----] A very important issue and we are struggling to get this up in a cost-effective way as research funds never pay for this type of activity.
- [-----] In a large, distributed and complex beast like [this university], the diversity of practice across the wide range of research disciplines means that evolving the infrastructure needed to support research in the digital era is not going to be easy. It is, however, essential. We have to be prepared to make mistakes, to try things out and experiment. We have to be very conscious of the broader framework in which we are working and constantly try to reveal the deeper principles of practice. Consequently, we have to be careful not to limit our research record management practices to what current technology offers - as this will have changed during the life of the project. On the other hand we do have to use the latest technologies to the best of our abilities to bring increased productivity and services to researchers. We have just entered the 'Wright Brothers' phase of the Digital era.
- [-----] It would nice to have support and training for databases, even within my own group
- [-----] As a member of the [committee for this department], we have discussed the issue of data storage and access taking into account participant confidentiality and risk. Any form of training for [department] members would be useful, I think.
- [-----] There is a real problem with time for any training. I hardly have time to do the work coming across my desk at the moment. With faculty and university re-structuring, it is possible that staff workloads have increased significantly.
- [-----] Anxious to learn more about data manipulation but time is running out!
- [-----] We are uncertain of the project's continuation at this stage. I am only an honorary now.
- [-----] Backup systems are the most important thing at the moment to understand and implement. It is easy to buy (and then fill) terabytes of storage for a workstation. To make it secure is extremely difficult.
- [-----] I'm not actually sure what eResearch is.....
- [-----] I am already collaborating with the e-research centre
- [-----] It has relevance in my field of ecology, which generally has a low level of data sharing and integration between projects (i.e. poor data storage and longevity)
- [-----] Practising e-research - yes - with the assistance of others.
- [-----] Please can training be online, as trying to find time for workshops is impossible and I would like all my staff to do it but need to manage within various commercial programs and teaching responsibilities
- [-----] I think the best place to store and manage data is a database, but it needs to integrate well into scientific analysis software. I would prefer the database itself to have the functions, graphing, and statistical tools built in, but I'm not sure that it is available - MS Access comes close.
- [-----] A major problem at University is the slowness of the intranet which greatly affects back-up processes to servers. It is so slow that transferring 1GB (for example) can take >2 hours. This results in staff and students in my group only backing up some files, with the risk of data loss.
- [-----] No time to think about it because of the very heavy workload
- [-----] I FEEL THAT SURVEYS LIKE THIS WASTE THE TIME OF RESEARCHERS. LESS BUREACRACY IS THE ANSWER.
- [-----] We have experts in our department on this.
- [-----] This has the feel of yet another make-work exercise that will ultimately get in the way of actually getting any research done.
- [-----] we have previously discussed data management and distribution with a university wide group [...] to establish data base format/archiving/etc standards.
- [-----] Data repositories and ethical data management are my specialties
- [-----] Not sure what is meant by the tern 'eresearch'
- [-----] re question 17 above, it would depend on the aims and time commitment required
- [-----] Re Q16, depends on how eResearch is defined.
- [-----] It is difficult to decide what is being offered - I don't want to sit through some long training program on how to manage huge digital data sets that are nothing like my own comparatively small sets.
- [-----] A key problem is discovering what solutions are available and who offers them for data management, digitisation, data storage/access in particular.
- [-----] Er, sorry, not sure what eResearch is? Data obtained through Web pages?

## General comments on data management

[-----] All of this sort of stuff is actually peripheral to doing the research - i.e. it takes time away from the actual stuff that you are meant to be doing.

[-----] I have not any idea about this question

[-----] Re questions 16 & 17, I consider myself as "coping" with the various forms of data management, certainly not as an exponent of eResearch, and certainly not able to provide any useful advice on what to do or not do!

[-----] test

[-----] It would help if there was assistance from [university] centrally in data management policy, guidance, infrastructure etc. It probably exists, but is seemingly inaccessible to individual researchers.

[-----] Needed area of consideration as I have only just started to realise that I need to be more organised in storing my data as my computer can not cope. Also being off site from [university] collecting data and analysing it raises issues with storage as well.

[-----] I realise I don't really know what I'm doing because this survey has raised a lot of questions that I hadn't thought about. I would certainly get a lot out of training. I'm guessing the situation would be similar for most students

[-----] Thank you for the survey. Sorry I can't join in the eResearch reference group (not the best moment in time, otherwise I would have done it with great pleasure).

[-----] Could have a "tree like" depository system where users could store them after the project ended.

[-----] PhD orientation program should cover this issue.

[-----] currently, student appear to be expected to manage data themselves. seeing other students, i think it is pretty poor where students that are funded by external parties are still expect to manage data themselves, rather than being given dedicated storage provided by the university or the project. DVD, flash drive or local hard drives is not really an acceptable storage and retrieval medium. it is also very very poor that it is now 2007, and the university as a whole still does not have proper electronic data management practices in place. i am aware of several projects over the past few years where the proverbial all data was lost because the university did not provide adequate systems to the student. the same philsophy shoudl also be used for undergraduate students. from what i have seen, providing student with a token amount of storage space just isn't really acceptable, especially when services such as google provide multi-gigabyte storage for no charge. it is also quite offensive when academic tell student to use these external services for data management as if to say to the student that data manage shoudl be in place, but the university isn't goign to provide any resources to the student to assist this. basically the universities in general are failing in their duty of care to the the students and certainly from a marketing perspective, grossly failing to provide adequate customer care their customers, being the students, internal funding bodies, and external funding bodies.

[-----] As far as I know, there is no secure area for me to store digital data for my thesis that is accessible anywhere in the world (e.g. via secure websites), has restricted access, and is backed up by the uni. This would be really useful.

[-----] This has the smell of yet another [university] bureaucracy that serves the group of geeks that run it but is nothing but an impediment for the researcher at the coal face

[-----] In Science there has been an almost complete refusal to acknowledge this issue. The policy (such that it is) is that the individual is responsible to archive and secure digital data. This covers research and teaching materials. I might add that server space is only made available on request and is subject to an unidentified process. I resorted to purchasing a back-up system from research funds, which transpired to be incompatible with the [university's system]. A complete waste of funds!

[-----] The amount of video data that we generate in our research will become increasingly problematic to store using current methods as I anticipate we will produce significantly more than 0.5TB a year of raw video through staff and postgraduate student projects. This does not include video transformed into formats suitable for analysis, reports and other materials generated after and through analysis.

[-----] What about the on line software survey project currently underway - [refer to ...]

## General comments on data management

[-----] Guaranteed integrity of all data is crucial for any research activity. Any server repository MUST maintain mandatory access profiling to prevent accidental or deliberate modification or destruction of stored data, e.g. malware (viruses, rootkits, etc). Any centralised service MUST demonstrate state-of-the-art backup and recovery facilities.

[-----] Some related case studies on intellectual capital and knowledge management were presented this week at this conference fyi: <http://www.academic-conferences.org/icickm/icickm2007/icickm07-timetable.htm>

[-----] It would be great if [this university] could offer its own online survey system. This would overcome concerns about an outside body having access to the data and would be useful to have good backup in case of problems.

[-----] I have pointed out to [this university's] staff that their data management was non-existent before [a system] was established. However, their asset, HR, student, IT, and financial systems seemed to take priority.

[-----] It is critical [this university] facilitates data intensive research with the provision of analytical infrastructure. The issue of curating completed project's data becomes much simpler if well documented datasets are established and utilised for analysis during the project. [This university] needs to foster SIGs/forums etc for the linking and amplifying the expertise that is diffused throughout faculties and projects..

[-----] Researching new media (VOIP, Video conferencing etc), need guidance on management, manipulation and storage to protect the data collected as well as commercial-in-confidence information.

[-----] Just a general comment on what I see with respect to storage of data in the research environment at [university] in general. Far too much data is stored in one location on one PC with one hard drive inside it. If that drive dies, which they do often enough, the data is lost forever. There's data on PCs so old they can not be networked for the purposes of backing up, and in some cases have no USB ports for extracting data, leaving just the floppy drive! The importance of off-site backup is often neglected. For example, having data stored on a laptop and a USB stick stored in the same backpack is of no use if the backpack is stolen!

[-----] 'Data management' has particular meaning in research - verification, consistency checking, cleaning, which are all post-collection issues and require their own specific protocol/plan. These are where resources are most needed by the typical researcher, and most cost-effective is IT input into design and formatting of customised systems for data collection, cleaning, and formatting in readiness for analysis. Rarely available and so researchers make do with limited understanding of the enormity of the task and longer-term ramifications of just 'making do'.

[-----] The university is considering a central data repository, but this is only useful if the data placed there is well organised and is accompanied by metadata which includes the (instrumental) conditions under which it was obtained. For example, spectral or diffraction data tends to be specific to the instrument it was measured on.