Data Sharing, Backup and Archival Tools

Learning Objectives for the Module

By the end of the module, learners will be able to

- Describe mechanisms one can use to share research information
- Describe processes for backup that can be used by researchers
- Describe mechanisms suitable for archiving research data
- Explain how to use one data sharing tool for dissemination of research information/results

Sharing Data

Introduction

- **Research data** - factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings.
- A **research data set** constitutes a **systematic**, partial representation of the subject being investigated.
- Research data does not include: laboratory notebooks, preliminary analyses, and drafts of scientific papers, plans for future research, peer reviews, or personal communications with colleagues or physical objects (*e.g.* laboratory samples, strains of bacteria and test animals such as mice).
- **Access arrangements**: The regulatory, policy and procedural framework established by research institutions, research funding agencies and other partners involved, to determine the conditions of access to and use of research data.

ACCESS TO RESEARCH DATA

- Effective access to research data, in a responsible and efficient manner, is required. We can take full advantage of the new opportunities and benefits offered by ICT tools.
- Accessibility to research data has become an important condition in:
  - The good stewardship of the investment in factual information/research;
  - The creation of strong value chains of innovation;
  - The enhancement of value from international co-operation.

WAYS OF SHARING DATA

There are various ways to share research data, including:

- Depositing them with a specialist data centre, data archive or data bank
- Submitting them to a journal to support a publication
- Depositing them in an institutional repository
- Making them available online through a project or institutional website
- Making them available informally between researchers on a peer-to-peer basis

Each of these ways of sharing data has advantages and disadvantages:

- **Data centres** may not be able to accept all data submitted to them due to formats, relevant domains, …;
- **Institutional repositories** may not be able to afford long-term maintenance of data or support for more complex research data;
- **Websites** are often ephemeral with little sustainability.

- Approaches to data sharing may vary according to research environments and disciplines, due to the varying nature of data types and their characteristics.

WHY SHARE RESEARCH DATA (1)

More specifically, improved access to, and sharing of, data:

- Reinforces open scientific inquiry;
- Encourages diversity of analysis and opinion;
- Promotes new research;
- Makes possible the testing of new or alternative hypotheses and methods of data analysis;
- Supports studies on data collection methods and measurement;
- Facilitates the education of new researchers;
- Enables the exploration of topics not envisioned by the initial investigators;
- Permits the creation of new data sets when data from multiple sources are combined.
- Sharing and open access to research data not only helps to maximise the research potential of new digital technologies and networks, but provides greater returns from the investment in research.

WHY SHARE RESEARCH DATA (2)

- **Scientific integrity** – To verify results; uncover errors
  - Publishing research data and citing its location in published research papers allows others to replicate, validate or build upon your results
  - Openly sharing research data also encourages the improvement and validation of research methods and minimizes the need for data re-collection
- **Funders and government** – some expect data to be made available for others to use with as few restrictions as possible, and in a timely manner and linked to the associated publications
- **Journal publishers-** An increasing number of journal publishers require the sharing of associated data
  - e.g. Figshare (http://figshare.com/) - repository where users can make all of their research outputs available in a citable, shareable and discoverable manner, in support of papers published by Taylor & Francis.
- **Recognition and impact** - Others who re-use data and cite it in their own research help to spread the word about the research and increase its impact
- **Collaboration-** Data sharing may lead to new collaborations between data users and data creators. Sharing data can often lead to improvements such as corrections in the documentation, or combination or comparison of datasets leading to new information
- **Funding application advantage** - without data sharing plan, no funding e.g. National Research Fund (NRF)

MAJOR ISSUES INHERENT IN PROVIDING DATA ACCESS

- **Technological issues:** access to research data, and their optimum exploitation, requires appropriately designed technological infrastructure, agreement on interoperability (e.g. metadata standards), and effective data quality controls.
- **Institutional and managerial issues**: while increased accessibility is important to all science communities, the diversity of the scientific enterprise suggests that a variety of institutional models and tailored data management approaches are most effective in meeting the needs of researchers.

MAJOR ISSUES INHERENT IN PROVIDING DATA ACCESS

- **Financial and budgetary issues:** scientific data infrastructure requires continued and dedicated budgetary planning and appropriate financial support. The research data's access, management and preservation costs should not be an add-on or after-thought in research projects.
- **Legal and policy issues:** Intellectual property rights and the protection of privacy, directly affect data access and sharing practices, and must be fully taken into account in the design of data access arrangements.
- **Cultural and behavioural issues:** Appropriate educational (for researchers) and reward structures are a necessary component for promoting data access and sharing practices. For those who fund, produce, manage, and use research data.
- Individual **research organizations** and even **countries** will need to determine the appropriate balance between the costs of improved access to this data and the benefits that result from such access, within existing financial limitations.

BARRIERS TO SHARING DATA

- A huge amount of data ends up unpublished, unshared and essentially wasted –particularly for datasets that have clear scope for wider research use, decision- making, policy making and hold significant long- term value
- Tension between the pressure to make data more open **earlier** on and the real fear researchers have that if they do that others will reap the benefits from the hard work they've done
- Culture of "my" data
- "We intend to make a patent application, and must avoid prior disclosure."
- "Don't want to make locations of members of endangered species available to poachers."
- "The research data are confidential because of the arrangement my research group has made with the commercial partner sponsoring our research."
- "My data form part of a long-term study upon which my research group is entirely reliant for its on-going research publications and academic reputation. We only share this with trusted colleagues."

Backup of Research Data

Introduction (1)

- Researchers are required to keep clear and accurate records of the procedures followed and results obtained, including interim results.
- Data generated in the course of research must be kept securely in paper and/or an appropriate electronic format.
- Such data to be securely held for a period of ten years after the completion of a research project, unless otherwise specified by the research funder or sponsor.

Introduction (2)

- Generally, researchers do not receive formal training in data management practices. Their levels of expertise are a problem as they are learning on the job
- Few researchers, especially early career, think about the long-term preservation of their data
- The demands of publication output overwhelm long-term considerations of data preservation
- Therefore, there is a great need for more effective collaboration tools, as well as online spaces that support the volume of data generated and stored, and provide appropriate privacy and access controls.

Why the need to Manage Data

- There are significant risks with not managing research data effectively
  - Confused data
    - Arising from lack of documentation meaning that experiments may have to be repeated to make sense of results of those experiments previously undertaken
  - Loss of data
    - May not be possible to repeat data collection
    - Loss of potential, opportunity and impact

Data Management Plan

Data Management Plan (DMP)

- A data management plan is a formal and practical document developed at the start of a research project which outlines all aspects of the data, including:
  - The nature of your data
  - How it is organised and described
  - How it is shared with others
  - How it will be stored in the long-term
- Developing a data management plan helps to ensure the research data are accurate, complete, reliable, and secure both during and after completion of the research.
- Funding bodies increasingly require that grant applications include data management plans e.g. America's National Science Foundation (NSF) required a supplementary document of no more than two pages labelled data management plan.

Benefits of Data Management

- Saves time – being able to find things
- Reduces possibility of data loss through managed back-ups, storage and security processes
- Reduces errors e.g. due to badly described data or confusion between file versions
- Enables you and others to find and understand what you have done through the provision of descriptions, metadata, file management etc.
- Provides evidence of work undertaken
- Provides evidence of validity of work undertaken
- Verifies – provides evidence of logical processes and methods
- Ensures retraceability and reproducibility for the research data

Data Protection, Backup and Archiving

- The terms data protection, data backups, data archiving, and data preservation have different meanings and purposes.
- **Data protection** covers a wide variety of topics including backups, archives, preservation, physical security (such as the use of smart cards), encryption, and others including laws which govern data security. However, this module will not focus on this broader topic but rather will look specifically at backups, archives and preservation.
- **Backups and Archives**: The terms data backups and data archiving are often used interchangeably as they both relate to saving a specific version of a file, but they are different.
- The term "**backup**" is used specifically when making copies of various files with the knowledge that the files may change.
- Backups are kept for a certain amount of time, but can be discarded after a specified time has passed.
- **Archiving** is used when a file is to be preserved as-is, often at the end of a project and acts as a static (and usually final) record.
- **Data preservation** encompasses many of these same methodologies, but can also include things like data rescue, reformatting of files, converting data, and the creation of metadata.

Backups vs. Archiving

- **Backups**
  - Used to take periodic snapshots of data at various moments in time to allow the user to restore the file as needed in case the current version is corrupted, destroyed, altered or lost.
  - Backups are copies of files stored for short or near-long-term
  - Backups are often performed on a somewhat frequent schedule
  - Backups are overwritten again as the data changes
- **Archiving**
  - Used to preserve data that is no longer in use for historical reference or potentially during disasters
  - Archives are usually the final version, stored for long-term, and generally not copied over
  - Archiving is usually performed at the end of a project or during major milestones or when appropriate according to procedures
- It is a good idea to have multiple copies of your backups and archives, in case one copy fails.

Reason for performing Backups

- Mitigate or prohibit the loss of data, which may or may not be reproducible
- Save time, money, and productivity as little to none of the data will have to be reproduced
- Having a backup already in place means you are prepared for when the unexpected happens, such as human error, disasters, or computer failures
- It allows you to go back to earlier versions and see what your results were. For example, if you are creating models and used data from an earlier model run, the most recent file you have on your computer may not have the same data as when you first created the model output.
- Backups provide for the ability to send older files to others, regardless of the current version or state (for example, if the current version has been corrupted)
- May allow you to respond during times when questioned results were based on older versions of files. For example, you may find that you will have to justify your results in court or to other scientists. By having access to older files, you may be able to respond to their requests for information. Or when you are not be able to reproduce the data, and the original copy may be the only evidence of the data collection.

Considerations for Backups (1)

- **Existing policies**: Your office or project may have existing policies on when, where, and how your backups can be performed. They may already have a backup procedure in place that you can use or build upon.
  - Backup policies may differ among groups: for example, your office may perform backups once a month, but your project may need to have the data backed up more often. In this case, you have to decide which one takes precedence
  - Policies between research groups may differ as to where the data backups live, and may have different restrictions on accessing the backups. Which one takes precedence?
- **Responsibility**: Data backups are often a small part of a good and comprehensive data management plan. Each data management plan should have specific guidelines on backups including when backups are performed, who is responsible for them, how the backups are accessed, where are they located. By having the answers to these questions, you are better able to manage your data and know who is responsible for various components related to data backups.

- **Confirm its done:** Many offices already have backup systems in place, managed by IT staff. But before you assume backups are being performed for you, you should confirm someone is responsible. Even if there is a backup system available, it may only cover certain enterprise-wide systems (like servers), and not desktops. And many offices have little or no IT staff, so performing backups may be solely up to you.
- **Fit for Needs:** You should identify and review the various policies (if available) and ensure they fit your needs and requirements. If they don't, you may need to discuss this with those involved in managing the backups or, establish your own schedule and plan.

Considerations for Backups (2)

- **Frequency**: How often do you want to do backups? Continually, daily, weekly or monthly?
  - The amount of time between backups depends on several things such as: can you afford to lose weeks worth of data if you only perform backups once a month? If not, then you should consider doing them more often. Are you creating real-time data that cannot be reproduced? If so, you'll want to consider continual backups.
- **Kind of backups:** Partial ones that only back up the data/files that have changed since the last backup and full backups which backup everything.
  - **Full backups** are required when beginning to establish your backups. They act as a full copy of all of your data. Then, **incremental (partial) backups** can be performed which will then backup any data that has changed since the last backup performed. Hence partial backups are often quicker and require fewer resources from your computer both in terms of processing and space.
  - It is still good to do full backups on a regular basis in case a previous full backup is inaccessible or unusable.
  - Do not overwrite the copies of your full backups. If you were to overwrite a prior copy of a full backup, you may find that the newer copy doesn't contain the same set of files as the older backup, or, the new file may be corrupted and then you are left without a viable full backup.
  - Hence, how often and what kind of backups you have depends upon how important your data is and what resources you have available.

Considerations for Backups (3)

- **Cost Vs Benefits:** If you are only occasionally working on a machine and the data isn't that important, then you probably don't need a top-of-the-range computer and backup software system which can be costly. Alternatively, if you are creating files for a multi-million dollar project, you don't want to be backing up your data by hand to an external drive.
- **Non-electronic files:** You should also think about how non-electronic files are backed up. A disaster damages all files – not just electronic ones.
  - Consider digitizing non-digital files so that they can be managed by an electronic backup system. Scan in the non-digital files at a high resolution so that you do not lose any information.
  - Even if the information contained in a non-digital file (such as a field notebook) is entered into an electronic system, you may still want to scan in the paper copy so that the format and presentation of the original file is preserved.

Considerations (4): WHERE TO BACKUP

Depends on a variety of things:

- Your office or project may have a specific location for where they want the backups to live
- In case of no system in place, consider using external drives, online centralized storage such as Dropbox, or using a pre-existing data repository such as GEON, NEON, GCMD or KNB, or using cloud services such as Amazon's or Google's Google Drive.
- While CDs and DVDs are cheap and frequently used to copy and share data, they have limited shelf-life and therefore are not reliable.
- Some repositories may require some form of metadata in order for you to upload your data. Be creating metadata records during the lifecycle of your project
- Even if you already have one backup in place, you may want to make sure your data is backed up in another location. In case of incidents such as fire or failure of the system.

Considerations

- **How** are backups carried out?
  - Manually may work for single files, but requires that the user remembers to perform regular backups and can be time-consuming
  - Automatic backups can be set to run on a set schedule that doesn't require the user to remember, and when having many files.
  - Many computers come with their own backup software, as do external hard drives.
- **What** do I do if I need to get a **file off backups**?
  - You should know how to obtain files from backups , where they are located (disk on local computer or centralized resource) and who to contact and how to contact the person if the backup system is handled by someone else, such as IT staff. There may be restrictions on who can access the files
  - Are the files backed up individually or as one large file?
  - You need to know this information beforehand, as often you need a file off a backup in an emergency! Make sure there is one person who knows how to retrieve the data file

Considerations

- How do you verify a backup has been successfully performed?
  - Most backup software will have a log file that contains details of the backup (which files, when the backup was created)
    - Good starting place
  - However, don't rely solely on the log file
    - Even if a log file states the backup was successful, you still need to check the backup to make sure the files are there, accessible and viable
    - Can you pull a file off of a backup and restore it to another location? Ensure this is possible
    - Hardware and software failures can happen at any time after backups are made and log files are created. As a result, a file might become unrecoverable
    - Your system might be backing up the wrong files and still reporting a successful backup!

Considerations

- How do you verify a backup has been successfully performed? (Cont.)
  - Since manual checks of all of the files in your backup is probably not possible, you should utilize other methods such as checking and comparing file sizes, date stamps, **checksum** values.
    - Checksum are mathematical calculations based upon a specific file. You use a program to calculate the checksum on the original file, then calculate the checksum on the backup copy.
    - If the calculated checksums match between the backup copy and the original file, chances are the original file is the same as the backup file and therefore was not modified when copied or stored.
    - There are various programs available to calculate checksums on files.

Archiving Research Data

Why archive your data?

Definition:

Archiving is used when a file is to be preserved as-is, often at the end of a project and acts as a static (and usually final) record.

Why archive research data:

- So you can continue to access and understand your data in the future because it is difficult to remember in detail what one was working on after a few years
- To prevent loss or inaccessibility of valuable knowledge and data when funding expires or people move on (to create memory of the research i.e. capture the knowledge stored in peoples' heads)
- So you can retrieve and share data easily if requested – to avoid a day searching for data and preparing it so it can be shared
- To allow data to be shared and to be combined in new and innovative ways

Considerations

- Are there backups of the backups? i.e. multiple versions of your backups that will also be on different media types and formats in case the primary backup fails
  - Necessary for high-value data
  - Usually different copies of backups are kept in different locations. You may also keep the current backup onsite for a week, but keep the previous three backups offsite, rotating the backups as new ones are made.
- How long do you keep your backups? - Keep full backups for a month, but with incremental backups you may need to do them once a week
  - Depends upon specific situation, but should be at least weeks or months.
- What happens to the backups (archives) after the project is no longer funded, project ends, or staff departs? Will your office or program take ownership? Look at agreements made earlier
- Which long-term storage solutions to use? Will data be archived elsewhere?

Considerations

- Can you read data off older backups?
  - Data storage media changes and you may no longer be able to read older versions and formats such as floppy disks, Jazz and Zip drives, Wordperfect files, etc. Older hardware e.g. floppy disks cannot be read by new PCs as they do not have readers for them. Older software and older file formats may no be read due to newer versions of the same program, which can lead to recovery failure
  - When new software versions come out, early on you can usually use the software to convert the older files into newer versions but not after a lot of time
  - When media readers e.g. floppy drives start disappearing, read the files using disappearing hardware and store them in new media e.g. flash disks
  - Even currently-available media (external drives and flash disks) are not immune to degradation. Media can degrade quickly, unexpectedly and inconsistently

- Hence, even if you can open a file today, that doesn't mean you can in a month from now
- Therefore, there is need to check the recoverability of backups on a regular basis

Considerations

- You will also want to consider what will happen to older backups.
  - Do you want to keep storing them as backups? i.e. copy them over to a new backup with other files
  - Should they be archived?
  - Should they be destroyed, and if so, how will you handle that?
  - If you are dealing with sensitive information, make sure you are using a reliable system that will completely destroy old files. Simply deleting a file off a computer or reformatting a hard drive does not completely prevent someone from accessing that information again. Special software may be needed to accomplish permanent deletion.

- Remember: only backup the data you can't afford to lose. For many of us, that is the majority of our files.

Considerations for data archiving and reuse

- Can you choose standards / formats etc that are more sustainable? Bear the long-term in mind when you're making these decisions. It's hard to backtrack and amend later.

– What do you want people to be able to do with the research data you are generating? Pick the right formats to allow the future use and preservation that you anticipate

– What information will future users need to understand the data – how will you make sure this information is captured? Plan and develop metadata and documentation from the start

– Is there somewhere you can archive the data? If so, do they have requirements / minimum standards you need to meet? Be aware of archive requirements so data is created to meet them

Metadata and Documentation (1)

Metadata Definition

- Metadata is...data about data. It is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource.

Metadata enables:

- Resource discovery and retrieval
- Data sharing and reuse – metadata allows data to be interpreted or analyzed by others
- Management of resources – metadata records aspects of the production and preservation process, rights information, location and access information

Metadata and Documentation (2)

Three broad categories of metadata are:

- **Descriptive** – consists of common fields such as title, author, abstract, keywords which help users to discover online sources through searching and browsing.
- **Administrative** - preservation, access rights management, and technical metadata about formats.
- **Structural** - how different components of a set of associated data relate to one another, such as a schema describing relationships between tables in a database.

What information is needed to interpret the data?

- Descriptions of all variables / fields and their values
- Code labels, classification schema, abbreviations list
- Details about how the data were created, analysed, anonymized
- Information about the project and data creators
- Tips on usage of the data e.g. exceptions, peculiarities, questionable results

Metadata and Documentation (3)

Considerations for choosing a standard for Metadata

- Whether it applies to the discipline/ domain
- Whether it supports the format of the data
- Repository or funder requirements
- Recognition and/or certification of standard
- Available metadata tools and whether they support the standard
- Skills required and time available to master and apply the standard

Examples of Metadata standards: Dublin Core, Irish Social Science Data Archive (ISSDA), Data Documentation Initiative (DDI), Common European Research Information Format (CERIF), Minimum Information for Biological and Biomedical Investigations (MIBBI)

Examples of metadata creation tools: MetaCenter, Morpho, Irish Social Science Data Archive (ISSDA) Data Deposit Form, Earthchem data submission form

Appraisal of What to Archive (1)

- Appraisal involves measuring the drivers for retaining a dataset or record against the costs of doing so, and determining the point at which the costs outweigh the drivers. It requires assessing the data against criteria such as:
  - Does the data or record fit into a repository's selection policy? Is there a selection policy in place at all? – find out first
  - Who will or might use the data or record in the future? Is there a defined 'designated community'?
  - Is it economically feasible to keep the data or record? What are the cost considerations for long-term maintenance of the data?
  - Can acceptable legal and intellectual property rights be negotiated to keep and re-use the data?
  - Is there a legal requirement to keep the data (and make it accessible) for a certain period of time?
  - Does the data constitute the 'vital records' of a project, organization or consortium and therefore need to be retained indefinitely?
  - Is it both technically feasible and worthwhile in cost/benefit terms to preserve the data or record? (What file formats are used, for example? Is their maintenance viable?)
  - Does sufficient documentation and metadata exist to explain the character, and enable the discovery of the data or record?

Considerations for data selection for Archiving (2)

- How significant are the data for research?
- How significant is the source in terms of scientific progress and society?
- Is the information unique?
- How usable are the data?
- What is the timeframe covered by the information? Does it apply to the next 10 or 20 years? If yes, archive
- Are the data related to other data in the archives?
- What is the volume of data? If it is a lot, then consider other factors otherwise, avoid archiving

http://www.dcc.ac.uk/resources/briefing-papers/introduction-curation/appraisal-and-selection

Best Practices

- Minimize or remove reliance on users to perform own manual backups (if possible reduce reliance on manual backups)
  - Implement standardized and automatic backups
  - If possible, put experts in charge of this task (computer staff) as they are more likely to keep up-to-date regarding software updates, hardware issues, best practices, etc.
- Don't assume backups are being performed for you
  - You don't want to find out after failure that no backups have been performed
  - If you are using third-party software (like Yahoo or Google Mail), what happens if they lose your files?
- Use non-proprietary, standard formats
  - Convert text files from .doc or .xls to .txt, image files to .tiff, or .pdf

Best Practices

- Check your backups manually
  - Start with log files, as they may tell you the backup was unsuccessful
  - Do not rely solely on the log files – they may be incorrect or the data may have become corrupted after the file was transferred
  - Look at file dates and file sizes to see if they match; calculate a checksum on the original and archived file and make sure they match
  - Ensure you can read files off older backups and archives.
- Have multiple versions of backups in multiple formats in multiple places/locations
- Good data management will limit the amount of data rescue that needs to be done to older data

An Example of Data Sharing and Backup Tool: Dropbox

What is Dropbox?

- Dropbox is a web-based service that syncs files across multiple platforms, regardless of filetype. But you're not limited to the web — there's a desktop application for PC, Mac and Linux AND apps for iPhone, iPad, Blackberry and Android.
- Dropbox allows you to save a file on your home computer, open it and edit it at work or school, and view it on your smartphone.
- Dropbox also automatically creates backup versions of your files.
- Basic Dropbox is free. If you need more storage space, you can use he paid versions of Dropbox

File syncing

- Edit and save a file on one device, and it automatically syncs to your other Dropbox-enabled devices.

File sharing

- Any file you save to the Public folder is assigned a public link you can share with others
- You can also invite specific people to share any other folder in your Dropbox folder

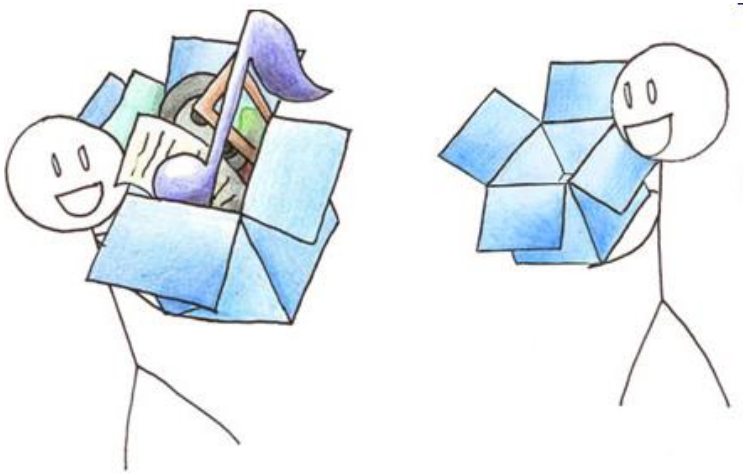
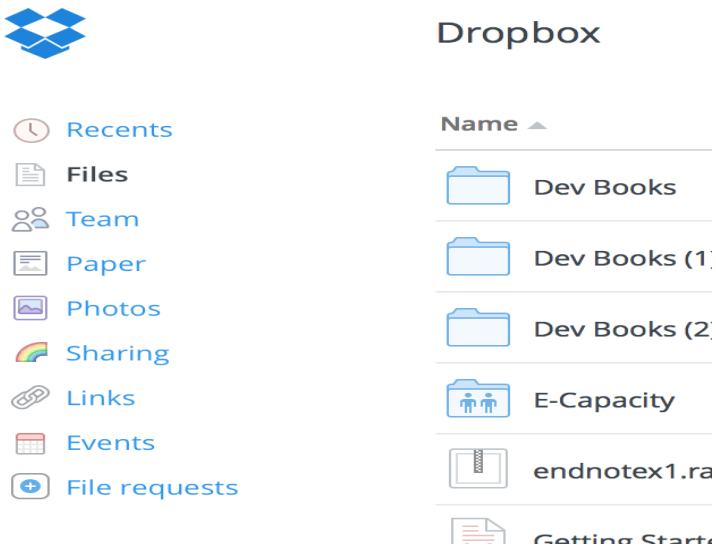


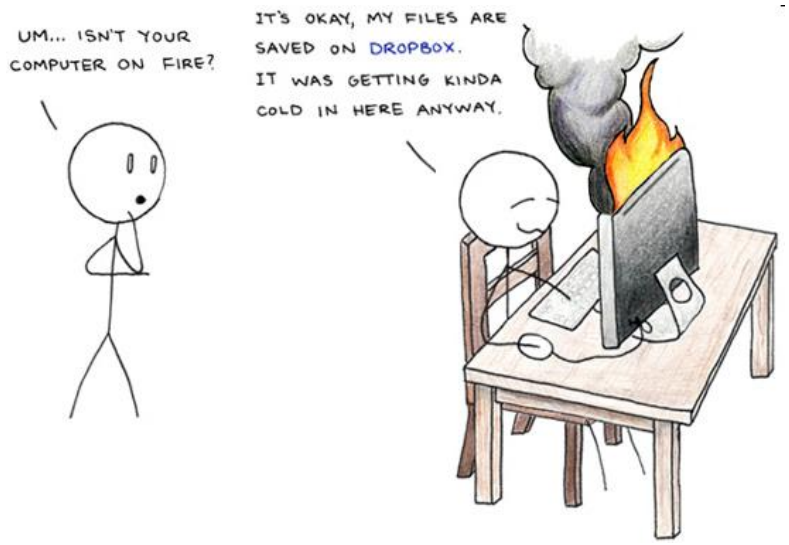
Photo sharing

- Move folders of images to your Photos folder to create galleries you can share easily with a public gallery link



Online backup

- When you save a file to your Dropbox folder, it's automatically backed up to Dropbox's servers.



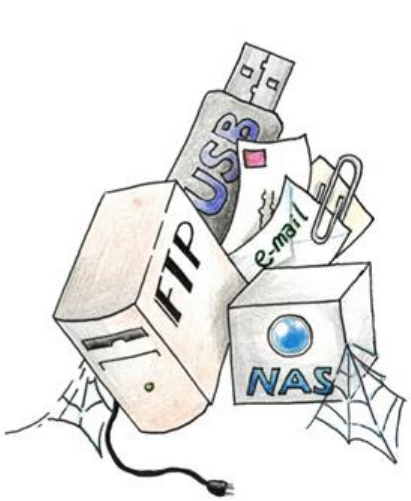
What it replaces

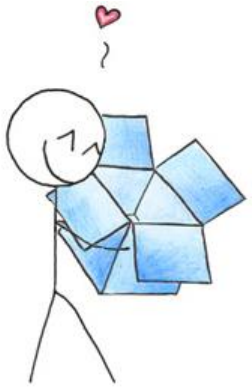
Saving multiple versions

External hard drives

Emailing files to yourself

USB drives





References

1. DataONE Education Module: Data Protection Backups. DataONE. Retrieved Nov12, 2012. From http://www.dataone.org/sites/all/documents/L06_DataProtectionBackups.pptx
2. *Backup*, wikipedia.org, http://en.wikipedia.org/wiki/Backup , (accessed 3/16/2011)
3. Georgia Tech Library, *NSF Data Management Plans – Research Data Management (*Georgia Tech Library and Information Center), http://libguides.gatech.edu/content.php?pid=123776&sid=1514980 (accessed 3/16/2011)
4. Albanesius, Chloe, *Google: Storage software update led to e-mail bug*, http://www.pcmag.com/article2/0,2817,2381168,00.asp (accessed 11/18/2011)
5. Van den Eynden, Veerle, Corti, Louise, Woollard, Matthew, Bishop, Libby and Horton, Laurence, Managing and Sharing Data, http://www.data-archive.ac.uk/media/2894/managingsharing.pdf (accessed 4/25/12)

For more information about physical security, encryption, and data disposal, visit: http://www.data-archive.ac.uk/media/2894/managingsharing.pdf

References

- http://www.oecd.org/science/sci-tech/38500813.pdf

OECD Principles and Guidelines for Access to Research Data from Public Funding

- Gavin Cole (). Archiving and Data Protection. The sun
- DCC data management and sharing plan guidance: http://www.dcc.ac.uk/resources/policy-and-legal/data-management-plans

- JISC briefing paper on digital preservation: http://www.jisc.ac.uk/media/documents/publications/digitalpreservationbp.pdf

- THE supplement, The data revolution: http://www.nxtbook.com/nxteu/tsl/jisc/#/0

- UKDA managing and sharing data guide: http://www.data-archive.ac.uk/news/publications/managingsharing.pdf

Key message: In order to archive your data, ideally you should plan and cost this in from the outset

- University of Glasgow. Archiving your research data
- http://www.dcc.ac.uk/resources/briefing-papers/introduction-curation/appraisal-and-selection
- https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwjNsvK3oObNAhWE2hoKHbhsD_EQFggcMAA&url=http%3A%2F%2Fwww.ucd.ie%2Ft4cms%2FJulia%2520Barrett-Managing%2520Data%2520.pptx&usg=AFQjCNHpTkHB9jkumNQmxD_yzevd2qBIog&sig2=Zz1A0-GJdsWNI_764NhdeA&bvm=bv.126130881,d.bGs
- http://www.ucd.ie/t4cms/Presentation12.pdf
- http://www.youtube.com/watch?v=Lc82pxxRkMo
- Dr Oboko. ICT Management Tools Workshop.