CSCI 109

Chapter 8:
Ethics, Conducting Research & Backups

"THE THING ABOUT QUOTES ON THE INTERNET IS YOU CAN NOT CONFIRM THEIR VALIDITY"

— ABRAHAM LINCOLN
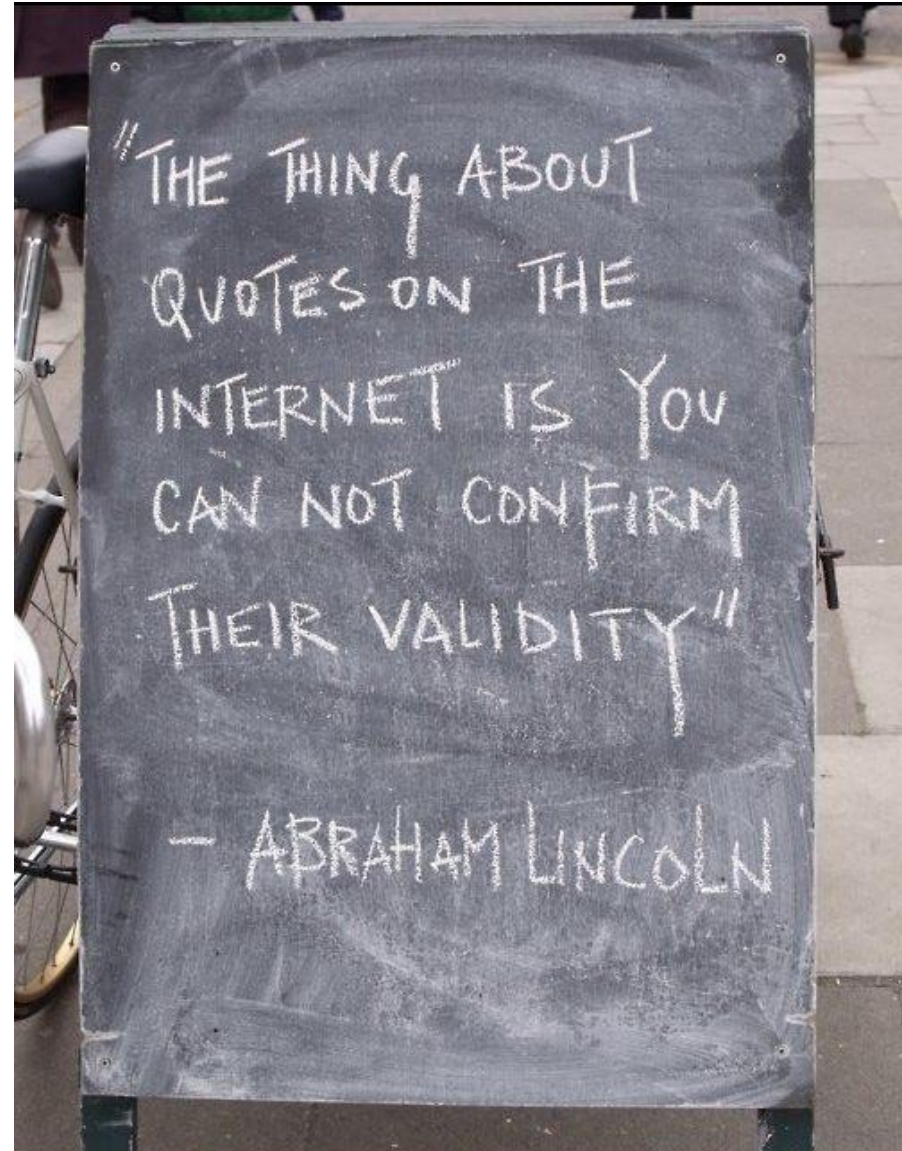
Topics

- Ethics and Computing
  - Plagiarism
- Conducting Research
  - Using the Internet for Research
  - Internet Research & Search Engines
  - Search Skills
  - Search Engines
  - Specialty Search Tools
  - Steps for Effective Searching
  - Do You Have to Study Every Resource?
  - Wikipedia
- Backups
  - Online Backups

Ethics & Computing

- Computer use
  - Ubiquitous in society
  - Universal tool

- Computer ethics
  - Principled choice between right and wrong in use of computer technology

Privacy

- Involves a freedom from intrusion, a right to be alone, and the right to control information about oneself

Freedom of Speech

- Organizations that lobby against censorship and restrictive legislation

  - American Civil Liberties Union (ACLU)

  - Electronic Frontier Foundation (EFF)

    - Blue Ribbon Campaign

Freedom of Speech

- A viewpoint:
  - Freedom must be balanced with providing protection against objectionable materials

- Child Online Protection Act of 1998 (COPA)
  - Effort by Congress to protect children
  - Challenged by the ACLU, EFF
  - Violated the First and Fifth Amendments to the Constitution
  - Law was not enforced; Supreme Court refused to hear the case in 2009, thus killing the bill

Intellectual Property

- Creations of the mind

- Means to protect owner's rights

  – Copyrights

  – Patents

  – Trademarks

Digital Material

- Easy to copy

- Challenging to enforce infringement
  – Example: Pirated music & movies

- RIAA and MPAA pursue copyright violations
  – Judgments, or fines, are awarded
  – Damages range from $750 per infringement to $150,000 ….per song downloaded!

Ten Commandments of Ethical Computing by Computer Ethics Institute

1. Thou shalt not use a computer to harm other people.

2. Thou shalt not interfere with other people's computer work.

3. Thou shalt not snoop around in other people's computer files.

4. Thou shalt not use a computer to steal.

5. Thou shalt not use a computer to bear false witness.

6. Thou shalt not copy or use proprietary software for which you have not paid.

7. Thou shalt not use other people's computer resources without authorization or proper compensation.

8. Thou shalt not appropriate other people's intellectual output.

9. Thou shalt think about the social consequences of the program you are writing or the system you are designing.

10. Thou shalt always use a computer in ways that ensure consideration and respect for your fellow humans.
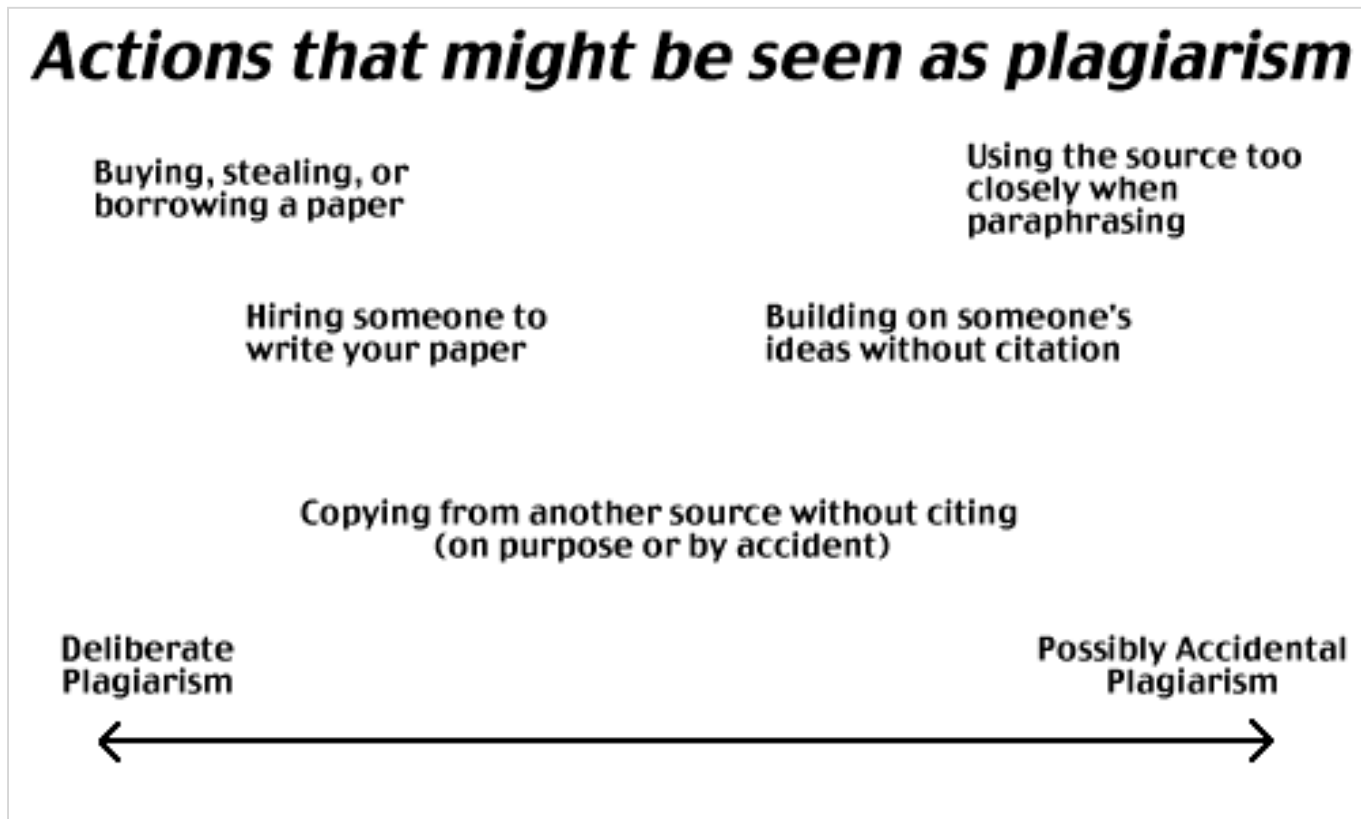
Plagiarism

- "Use (another's production) without crediting the source".
  - [Merriam-Webster]

- Software
  - **Plagiarism helpers** pull information directly from numerous sources on the web
  - **Plagiarism detection** can find matching text and give instructors links directly to the sources used

Misconceptions regarding information copied from internet

1. Everything on the Internet is free

2. Professor can't check every paper

3. It's not a big deal

4. Use plagiarism detection software before turning in

Last word on plagiarism

- Best approach is to do your own work and cite your sources

Actions that might be seen as plagiarism

Buying, stealing, or borrowing a paper

Using the source too closely when paraphrasing

Hiring someone to write your paper

Building on someone's ideas without citation

Copying from another source without citing
(on purpose or by accident)

Deliberate Plagiarism

Possibly Accidental Plagiarism

←——————————————————→

Conducting Research

- Finding useful, correct information on the Web can be difficult

- Search terms can have many meanings and interpretations

Hebridean Ram
Source: Wikimedia / L. Boynton

Ram Truck
Source: public domain image

Research:

- *A targeted and systematic attempt to discover facts, information supported by reasonable argument or consensus, reliable data, and even opinion (when recognized as such) in an attempt to further understanding*

- Scientific research is different
  - Research following a defined and rigorous process

Internet Research

- Print-only resources
  - physically limit access to information
  - Materials must be identified and then obtained

- Internet resources
  - Web can be searched
  - Thousands or even millions of pages in results

Google "cold fusion" physics

Search   About 754,000 results (0.12 seconds)

Search Results
Source: Constructed image

Evaluating Internet Authors

- Pertinence and accuracy of information
  - Subject to writer's expertise, credentials or credibility, presence of bias

How the Media can manipulate our viewpoint

Source: Kaifu Zhang

Search Engines

- Search engines display websites based on:
  - Popularity
  - Relevancy
  - Geographical Location

- Search results are subject to
  - Political bias
  - Social bias
  - Economic bias (advertisers)

  - Google Bombing – example of manipulating results

Search Skills

- Basic search queries include:
  - Phrases
  - Questions
  - Set of keywords
- Search engines give higher rankings to
  - Results with the query words in the document title or keywords, or near the top of the page
  - Results that are "popular"
- Basic search is a sufficient strategy for casual searching

Search Modifiers

- NOT, OR, AND
  - These are Boolean operators

- Advanced Search
  - Word stem
  - Dates
  - Quote
  - Form fill-in interface
  - Search previous results

Search Results

- Websites are created in formats dictated by
  - Tradition
  - Technology
  - Particular audience

Web crawlers

A web search engine used to index and sort the Web's contents -

1. Follows hyperlinks that connect pages to WWW

2. Information on Web pages is indexed and stored

3. Search results then ranked using complex algorithms
    - location and frequency of keywords on a Web page
    - quality and number of external hyperlinks pointing to Web page

Search Engines
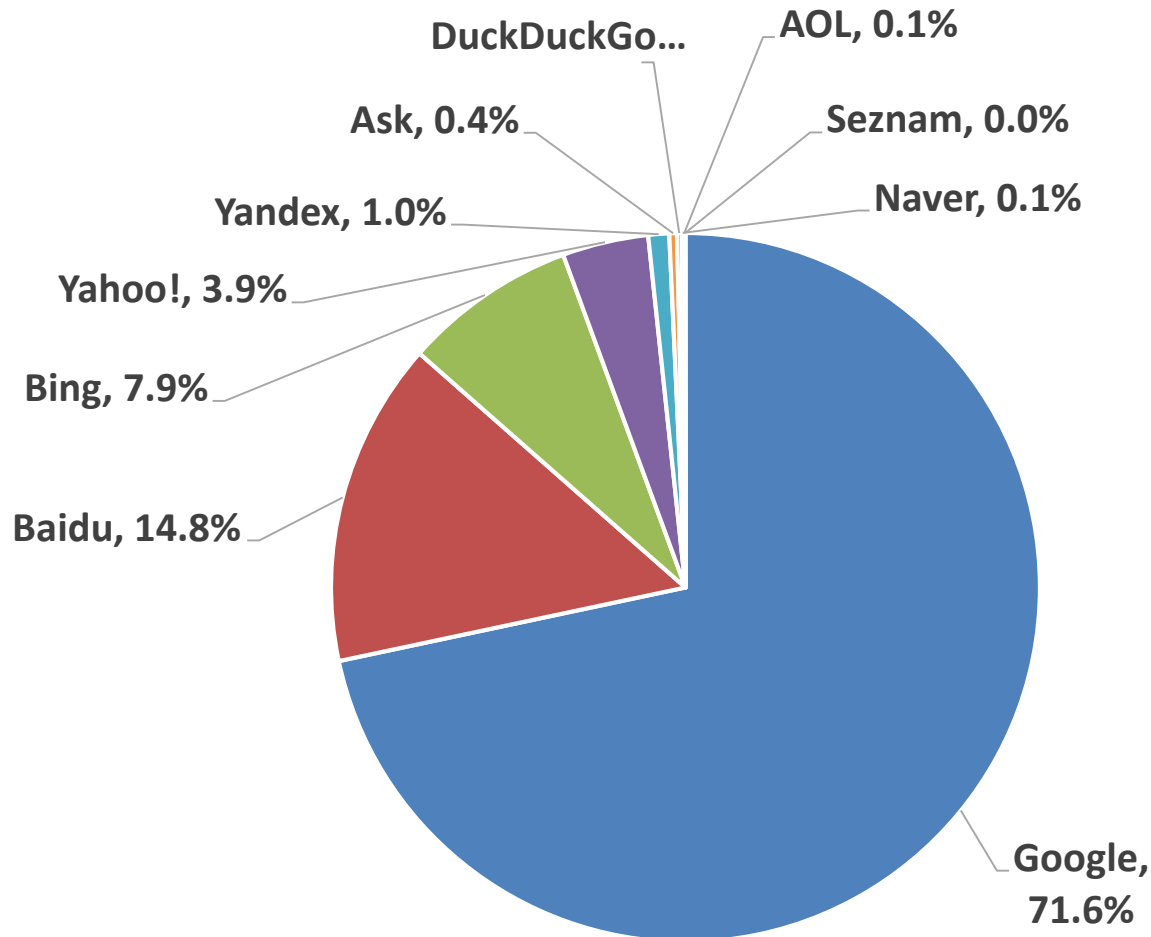
Which search engine do you use most often?
Desktop Search Engine Market Share, June 2018
Source: Netmarketshare.com

Metasearch Engine

- Runs search query against multiple search engines simultaneously

- Creates a list of aggregated search results

- Example: Dogpile

  – Combines results from Google, Yahoo & Bing
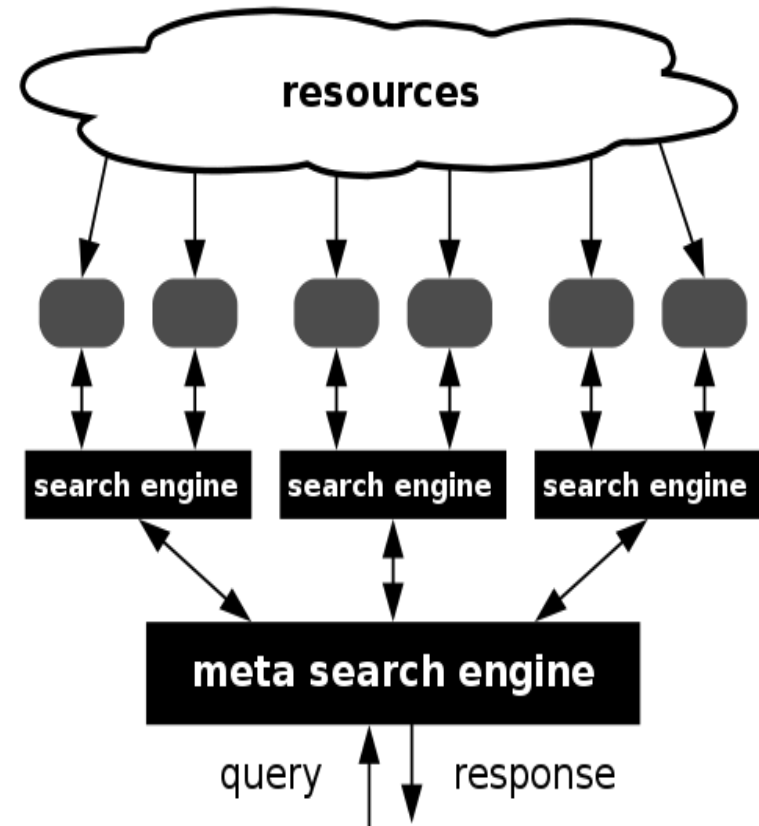
Diagram of the Metasearch concept

Source: Wikimedia / J. Voss

Web Directory

- Organizes subjects in an hierarchical fashion

- Let users investigate a specific topic

- Assembled by
  - Algorithms (more comprehensive coverage)
  - Hand crafted (more reliable)

- Examples:
  - Open Directory Project, Yahoo! and the WWW Virtual Library

Specialty Search Tools

- Enables users to find information stored in databases

- Provides users to quickly & easily
  - Find 'deep Web' content
  - Information that cannot be accessed by conventional search engines

- Relies on advanced bot & intelligent agent technologies

Internet Research Software

- Capture information while researching

- Collects information in one place for easy access

- Information is organized
  - Tagging
  - Hierarchical trees

- Examples: [Surfulater](), [Evernote](), and [WebResearch Professional]()

Steps for Effective Searching

1. Getting started

2. Planning

3. Digging for answers

4. Look before you click

5. Evaluate your sources

Getting started

- Determine topic and type of research
  - Personal information of for term paper?
  - Importance of accurate results?

- Best answers may not use Internet search
  - Library databases are a great resource
    - ERAU's Hunt Library
  - Research librarians can help immensely

Planning:

- If a web search is required, determine terms, tools and sources

- Generic terms vs. technical or professional terms

- Limit search to specific domains
  - Try it, in Google type: physics site:mit.edu

- Use online encyclopedias to topic overview
  - Are references at end usable primary sources?

Digging for answers

- Research may require many interrelated searches
  - Use multiple search engines
  - Go past the first few screen of results
  - Do not care for 'popular' answer

- Refine results
  - Use additional keywords to narrow results
  - Use "exact phrase" search with quotes
  - Expand search to non-English sites –use Google Translate

Look before you click

- The URL domain may provide clues
  - .com is commercial   /   .edu is educational
  - .org may or may not be a nonprofit

- Expertise may be hard to distinguish
  - Is a written by an expert on a topic of a fan club member (for example)?

- Does the domain indicate bias at any level?

- When was it written? Is the content current?

Evaluate your sources

- Who owns the domain?
  - Use a "whois" server like [Internic](#) to find out, then Google that organization or person

- Does the content reflect bias?
  - (e.g. New York Times, CNN, Fox, etc.)
  - Bias needs to be recognized, but doesn't make the information unusable.

- Is it satire, a parody, spam?

"judging a book by its cover"

- Short cuts students take:

  – Skip the detailed analysis and simply look at the web page

  – if it looks good, it must be a good source

    - not always a good call

Susan Boyle on [Britain's Got Talent](#). Ridiculed due to looks, received standing ovation.

Source: screen capture

Proper Source Evaluation

- When done properly, you are doing the equivalent of reviewing the following from a book:
  - copyright page
  - reading about the author
  - look at the index and table of contents
  - flip through the book looking for professionalism

Wikipedia

- Professors may not permit citing Wikipedia in a research paper
  - A Wikipedia article can be terribly wrong
  - As anyone can edit any page, vandals sometime make false edits

- Most professors won't mind students that *start* their research using Wikipedia
  - Most articles have a list of references and external sources at the bottom of the page

Wikipedia article sources

References

1. ^ The founder of Wikipedia is the sole individual empowered to override this process, but has stated in public that extreme circumstances aside, he will not do so.
2. ^ Jeff Young (June 12, 2006). "Wikipedia Founder Discourages Academic Use of His Creation". The Chronicle of Higher Education.
3. ^ Andrew Orlowski (28 May 2006). "New Age judge blasts Apple". The Register.
4. ^ Andrew Orlowski (15 June 2006). "Avoid Wikipedia, warns Wikipedia chief". The Register.
5. ^ Noam Cohen (29 January 2007). "Courts turn to Wikipedia, but selectively". The New York Times.

External links

- How to Evaluate a Wikipedia Article – A one-page PDF with similar recommendations to this page.
- Evaluating Web Pages: Techniques to Apply & Questions to Ask from the University of California, Berkeley
- Critically Analyzing Information Sources from Cornell University

Example of references and external links at the bottom of a Wikipedia article

Source Wikipedia

Backups

- Reasons to back-up data
  - Hard disk failure
  - Deleted files by accident
  - Virus
  - Lost/destroyed notebook
  - Data deleted off server
  - Natural influences like thunderstorms

Lost Laptop

Source: **Flickr**

Back-up Media

- Choose a backup method and backup media

- RAID: Redundant Array of Independent Disks
  - For businesses with complex data centers

- External hard drive/tape or optical disk

Full Backup

- All system data is copied to a backup media
  - Typically once a day

- In event of data loss
  - Most current backup media is copied back to the system´s hard disk

Backup Window

- Time frame available to perform backups

- Should be performed when the system is not used

- Window might be too small or data too large to perform a full backup every day
  - A challenge for full backups

Incremental Backup

- Full backup once a week

- Performs a backup every day that copies only files that have changed

- Backup window can be smaller than for a full backup
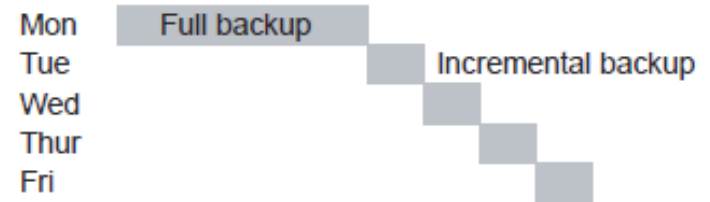

Incremental Backup

Source: Novell

Restoring Incremental Backup

- To restore the server

  - need all incremental backups and the full backup

  - All backups need to be copied to the server in the correct order

Differential Backup

- Performs a full backup once a week

- Performs a backup every day
  - records files that have changed since the last full backup

- Shorter average time needed to restore a system
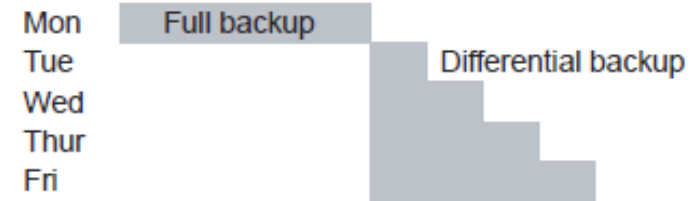
Differential Backups

Source: Novell

Differential backup restore

- To restore data, need two backup media:
  - the last full backup
  - the last differential backup

- Amount of data to be backed up grows every day
  - Amount of data might be too large for the available backup window at end of backup cycle

Backup Media

- Tape drives:
  - Best price-to-capacity ratio
  - Can be reused
- Removable hard drives
- Writable CDs or DVDs
- Magnetic-optical (MO) drives
  - Also rewritable
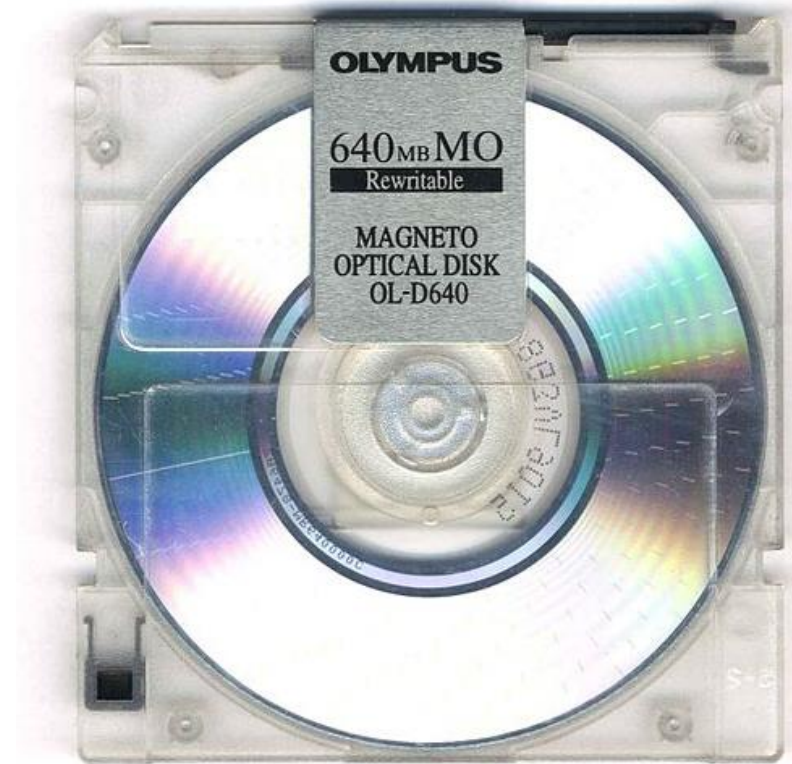
Magnetic-Optical Drive

Source: Wikimedia

Storage Area Networks (SANs)

- Backs up data from different computers on a central backup server

  - May use magnetic tapes or hard drives

- Backup media are stored separately from the backed up systems

  - Prevents backups from being lost

Online Backups

- Automated online backup requires:

  – Internet connection

  – Software program

  – Set up a backup schedule

  – Identify the files and folders to be copied.

How Online Backup Works

- Software sends copies of the files to a remote repository via the Internet

Online Backup

- Online backup is not equivalent to online file storage

- Automated online backup is ideal for small businesses
  - To store critical information: transactions, financial data
  - Lack the equipment or inclination to set up dedicated on-site storage.

- Online remote backup moves data to a third-party facility

Online backup: Pros

- The expense of purchasing backup equipment is not incurred

Hurricane Katrina

Source: NOAA

- Critical data can still be recovered in the event of a disaster

- Automation is another key benefit to remote backup

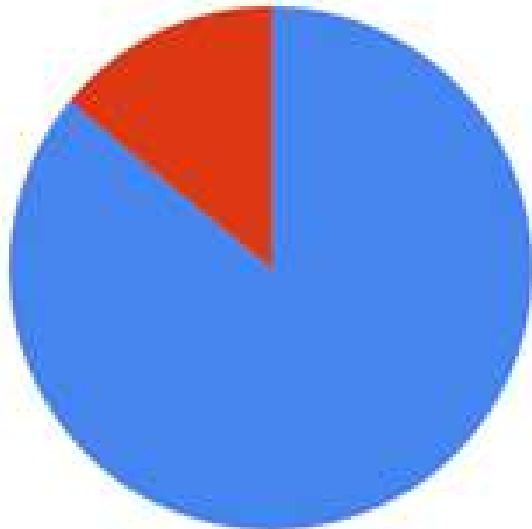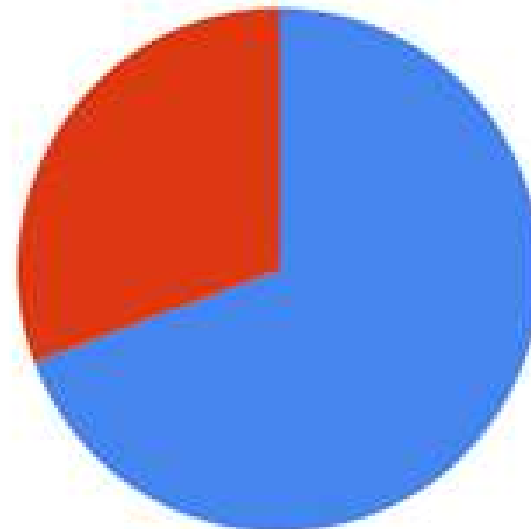# Regular backups and clear documentation go hand in hand

Does your organization back up its electronic records and other data on a regular basis?

In the last two years, has your organization taken steps to identify its mission-critical data and describe that data (e.g. how it's stored; who has access to it; how it's backed up)?

Yes
No

Regular backups and documentation go hand in hand.

Source: Techsoup

Online Backup Service Options

- Acronis
- Backblaze
- Barracuda Backup Service
- BullGuard Backup
- Carbonite
- Crashplan
- Dolly Drive
- Diino
- Dropbox
- Egnyte
- ElephantDrive
- Fabrik Ultimate
- FilesAnywhere
- Humyo
- IASO Backup
- IDrive
- Jungle Disk

- KeepVault
- Memopal
- MiMedia
- Mozy
- SpiderOak
- Syncplicity
- Ubuntu One
- Unitrends Vault2Cloud
- UpdateStar Online Backup
- Usenet backup
- Windows Live Mesh
- Wuala
- Zetta Enterprise-grade Online Backup
- Zmanda Cloud Backup
- ZumoDrive

How to evaluate?

- List out features important to you

- Refer to websites such as:

http://online-data-backup-review.toptenreviews.com/

Floppy Discs:
Not a Good Backup Idea

Source: **Flickr**

Last Slide

End