Data Storage Best Practices

BILL COREY

RESEARCH DATA MANAGEMENT LIBRARIAN

UVA LIBRARY

RESEARCH DATA SERVICES AND SCIENCES

Welcome!

Introduce yourself, and tell us what you would like to get out of this workshop.

Introduction

Data Storage is a critical part of every research project. This workshop will explore digital data storage best practices, storage options available from UVa through ITS, options from 3rd parties, data sharing, and data security requirements. We will look at UVaBox, Box, Dropbox, Google Docs, Google Drive, Amazon AWS, and Microsoft OneDrive, SharePoint, and Azure.

Let us start with what data is/are, followed by a look at the history of data storage. We will visit a few websites that show the timeline for several different data types.

What is/are Data?

data noun, plural in form but singular or plural in construction, often attributive

da·ta | \ˈdā-tə, ˈda- also ˈdä- \

Definition of data

1 : factual information (such as measurements or statistics) used as a basis for reasoning, discussion, or calculation

2 : information in digital form that can be transmitted or processed

3 : information output by a sensing device or organ that includes both useful and irrelevant or redundant information and must be processed to be meaningful

https://www.merriam-webster.com/dictionary/data

Data Types

Data generally fall into 5 categories:

- **Observational**: Captured in real-time. Cannot be reproduced or recaptured. Sometimes called 'unique data'. Example include sensor data, human observation, and survey results.
- **Experimental**: Data from lab equipment and under controlled conditions. Usually reproducible, but expensive to do so. Examples include gene sequences, chromatograms, spectroscopy.
- **Simulation**: Data generated from test models studying actual or theoretical systems. Models and metadata where the input may be of greater importance than the output. Examples include climate models, economic models, systems engineering.
- **Derived** or **Compiled**: The results of data analysis, or aggregated from multiple sources. Reproducible, but very expensive. Examples include text and data mining, compiled databases, 3D models.
- **Reference** or **Canonical**: Fixed or organic collection datasets, usually peer-reviewed, and often published and curated. Examples include gene sequence databanks, census data, chemical structures.

Data File Formats

A file format is the structure of how information is stored in a digital file. There are literally hundreds of file formats. Many are obsolete or proprietary. You want to use ones you are comfortable with, that are required by a specific instrument, process, or software program.

Non-proprietary, standards-based, and open formats are preferred for data sharing and data preservation. Unencrypted and uncompressed formats are recommended for images, video, and audio files. Proprietary formats are sometimes required to preserve the richness of data.

When you are working with data on a project it makes sense to use the formats that work best with the software and instruments you need. Therefore, Active storage should be format agnostic. When you have completed using that data file, you should think about the correct format for long-term storage.

Format obsolescence is a very real and serious problem. Data migration to newer formats and fresh media should be part of your data storage strategy.

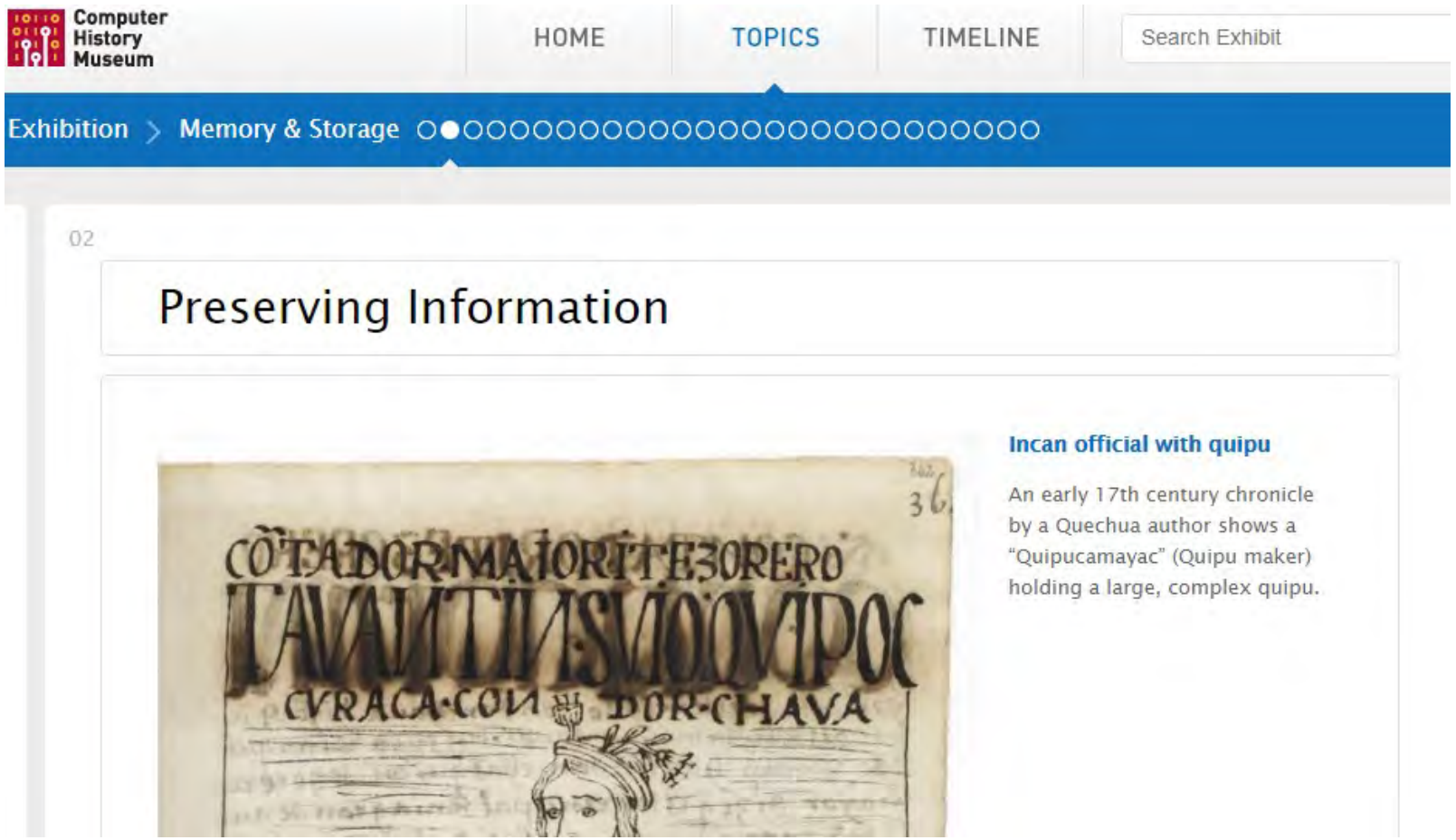
Recommended Digital Data Formats

- Text, Documentation, Scripts: XML, PDF/A, HTML, Plain Text.
- Still Image: TIFF, JPEG 2000, PNG, JPEG/JFIF, DNG (digital negative), BMP, GIF.
- Geospatial: Shapefile (SHP, DBF, SHX), GeoTIFF, NetCDF.
- Graphic Image: raster formats: TIFF, JPEG2000, PNG, JPEG/JFIF, DNG, BMP, GIF.
- Graphic Image: vector formats: Scalable vector graphics, AutoCAD Drawing Interchange Format, Encapsulated Postscripts, Shape files.
- Graphic Image: cartographic: Most complete data, GeoTIFF, GeoPDF, GeoJPEG2000, Shapefile.
- Audio: WAVE, AIFF, MP3, MXF, FLAC.
- Video: MOV, MPEG-4, AVI, MXF.
- Database: XML, CSV, TAB.

http://www.loc.gov/preservation/resources/rfs/index.html

History of Data Storage Devices

History of Data Storage Devices



History of Data Storage Devices

vodien Blog Topics About

Blog / Education / A Chronology of Computer History

A Chronology of Computer History

Updated on May 18, 2018 - by Val Soh

Modern computer technology offers a universe of conveniences. People can loc
online and even work from home via the Internet, while cutting edge technology
doctors to do surgery on people from around the world. This is due to the abunc
modern gadgets including laptops and tablets, as well as personal desktop com
speed Internet lines, and web hosting servers. In order to appreciate these techr
wonders, however, it is important to know how technology got to this point. It is
thousands of years of technological evolution and progress across many parts

3000 BC: The first abacus is likely invented and used by the Babylonians.

1800 BC: Algorithms are developed by a Babylonian mathematician as an aid in solving

500 BC: The Egyptians invent the bead and wire abacus.

200 AD: Computing trays were introduced by both China and Japan at this time. For China it was the saun-pan computing tray; in Japan the soroban computing tray was invented.

1000: Pope Sylvester II devised and introduced Eu
his knowledge of Hindu-Arabic numerals.

1617: The Scottish inventor John Napier demonst
numbering rods made of ivory. These rods becam
material they were made of.

1642: The first mechanical calculating machine, t
was 19 years old.

1673: A mechanical calculating machine that adds
by Gottfried Leibniz.

1805: Joseph Marie Jacquard invented punch car
mechanical loom.

1822: Charles Babbage proposes and designs a D
however, the machine is never completed.

1833: The Analytical Machine, which is considered
designed by Charles Babbage.

1842: Babbage's work is documented by Ada Lovelace. In addition, she
for him.

1854: The Mathematical Analysis of Logic is published by George Boo

1855: The first practical mechanical computer is based on work by Bab
George and Edvard Scheutz of Stockholm.

1884: A patent application for a punch-card tabulating machine is filed

1885: The first commercial listing and adding machine was developed
William Burroughs.

1889: The Hollerith tabulating machine patent is issued.

1890: To aid the taking of the U.S. census, Dr. Herman Hollerith creates
perforated cards.

1903: Electric logic circuits known as gates or switches are patented b

1924: International Business Machines becomes the new name for the
Tabulating-Recording Company.

1931: Konrad Zuse builds Z1, which is the first calculator.

1933: The Voder, a talking machine and the first of its kind, is invented by Homer Dudley for Bell Telephone Laboratory.

1937: The first binary calculator is created by George Stibitz for Bell Telephone Laboratories.

1938: Hewlett-Packard Co. is founded.

1939: Homer Dudley demonstrates his voice encoder called VOCODER at New York's World Fair.

1939: A prototype for the ABC (Atanasoff-Ber
and Clifford Berry.

1940: The first terminal is created as a result
Laboratories.

1941: Alan M. Turing designs the Colossus c
at the University of Manchester.

1941: A calculator with automated controls is

1944: Colossus Mark II is built in England.

1944: The relay-based computer called Mark
Hopper becomes its first programmer.

1946: Eckert and Mauchly start the first real-t
Computer), which takes three years to comple

1946: ENIAC (Electronic Numerical Integrator and Computer) is dedicated at the University of Pennsylvania.

1946: The Universal Automatic Computer Univac is designed by the Electronic Control Co., which was formerly Eckert-Mauchly Computer Corporation.

1946: John Tukey coins the word "bit" for binary digit.

1947: An Intelligent Machinery article by Alan M. Turing defines artificial intelligence.

1947: The not-for-profit Association for Computing Machinery (ACM) is founded.

1948: Maurice V. Wilkes of the University of Cambridge develops the Electronic Delay Storage Automatic Calculator, or EDSAC.

1948: The 12,000 tube Selective Sequence Electronic Calculator (SSEC) is created by IBM.

1948: William Bradford Shockley, John Bardeen and Walter H. Brattain invent the transistor.

1949: The first tests of magnetic disks are supported by EDVAC (Electronic Discrete Variable Automatic Computer).

1950: Assembler (symbolic assembly language) is applied to EDSAC by Cambridge University's Maurice V. Wilkes.

1950: The National Bureau of Standards receives SEAC (Standards Eastern Automatic Computer).

Data Format Timeline

Data Format Timeline

From Jacquard Loom cards to PlayStation Vita memory cards, the history of data storage on removable media. Comprising disks/discs, tape, solid-state media, ROM cartridges, and punched media.

Dates are approximate and refer to availability in the UK or Europe where known, otherwise for US or elsewhere.

1800s

Jacquard Loom card (1801 – 1990s)

1840s

Punched tape (1846 – 1980s)

1890s

Herman Hollerith's punched card design is used in the 1890 US census (1890)

Punched card (1890 – 1980s)

1940s

Aperture card (1943 – 2000s)

Open reel instrumentation and data logging tape (1949 – 2000s)

1950s

UNISERVO, the first magnetic tape drive for a commercial computer system is introduced (1951)

IBM introduces 7-track magnetic data tape (1952)

IBM introduces the hard disk drive with the IBM 305 RAMAC using 50 24-inch disks storing 5 MB (1956)

LEO tape (1958 – 1981)

Data Pages

Latest Data Formats

Leapster (2003 – 2012)
Sharp Pocket Disk (1986 – early 1990s)
LeapPad (1999 – 2008)
IBM 3490E (1991 – 2004)
Sony PlayStation Vita game card (2012 –)

Data Format Tags

½-inch ¼-inch 1292 Advanced Programmable Video System 12cm 1940s 1950s 1960s 1970s 1980s 1990s 2000s 2010s 3.5-inch 4mm 5.25-inch 80mm 8cm 8mm analogue Atari audio blue laser cartridge cassette CD-ROM Compact Disc disc disk DVD eighth generation endless loop fifth generation flash floppy fourth generation game console

Museum of Obsolete Media

Audio Format Timeline

Audio Format Timeline




A brief history of audio recording and playback, from the 1850s onward, including details of all the audio formats in the Museum.

Dates of individual formats are approximate and refer to availability in the UK or Europe where known, otherwise for US or elsewhere.

1850s

Frenchman Édouard-Léon Scott de Martinville uses the phonautogram to record the human voice by tracing sound waves on smoke-blackened paper or glass. The resulting tracings could not be played back at the time, but in 2008 several tracings from 1860 were processed as digital audio files and successfully played back (1853)

1870s

Thomas Alva Edison succeeds in recording and playing back 'Mary had a little lamb' on the first phonograph using tinfoil wrapped around a cylinder. He receives a patent in 1878 for recording on tinfoil (1877)

Organette disc (late 1870s – 1920s)

1880s

Piano roll (1883 – 2008)

Music box disc (1886 –)

Bell and Tainter are granted a patent for their graphophone, which uses wax-coated cardboard tubes instead of tinfoil, and engraves the sound waves instead of embossing them (1886)

Organ cobs (late 1880s – late 1920s)

Graphophone / Dictaphone cylinder (1887 – early 1950s)

Emile Berliner is granted a patent for gramophone discs (1887)

Edison introduces his 'Perfected Phonograph' using all-wax cylinders (1888)

Brown wax cylinder (late 1880s to 1906)

Ediphone (1888 – early 1950s)

Berliner's first gramophone discs (of 5 inches diameter) are marketed in Europe (1889)

1950s

Tefifon (1950 – 1960s)

SoundScriber tape (1950s – 1980s)

Flexi-disk (1950s –)

16⅔ rpm LP (Long Play) 12 inch record (early 1950 – early 1970s)

Minifon wire reel (1951 – 1967)

Cook Binaural record (1952 – 1958)

The 7-inch EP record is launched, sitting between the 7-inch single and the 12-inch LP (1952)

7-inch EP (1952 –)

In the UK, a record chart for sales of singles begins, initially with just the top 15 (1952)

Pye magnetic disc (1953 – late 1950s)

Seeburg Background Music Library (1954 – 1960s)

Grundig Stenorette (1954 – 1970s)

Mohawk Midgetape (1955 – early 1960s)

Chrysler starts putting Highway Hi-Fi players in its cars in an attempt to allow drivers their choice of music other than the radio, but the system is abandoned in 1959 (1956)

Highway Hi-Fi (1956 – 1959)

The first commercial stereophonic LPs are released (1957)

Unit sales of 78 rpm records reach a peak of 54.1 million, before quickly declining (1957)

Stereophonic LP (Long Play) 12 inch record (1957 –)

Dictaphone Dictet (1957 – early 1960s)

Philips EL 3581 (1958 – early 1960s)

RCA introduces the Sound Tape Cartridge, offering the sound quality of stereo open-reel tape, but in a much more convenient pre-threaded form. It lasts until 1964 (1958)

RCA Sound Tape Cartridge (1958-1964)

Minifon tape (1959 – 1967)

Seeburg Background Music System (1959 – 1986)

Fidelipac (1959 – late 1990s)

2000s

Compact Disc sales peak in US at 942.5 million units, and decline each year after (2000)

Copy-protected Compact Disc (2000 – 2006)

DVD-Audio (2000 -)

Apple launches the iPod (2001)

Timecode vinyl (2001 –)

e-kara Karaoke Cartridge (2001 – 2009)

VJ Starz Video Karaoke Machine (2002 – mid 2000s)

DataPlay (2002 – mid 2000s)

Record labels agreed to licence music to Apple to sell, and the iTunes Store is launched (2003)

Most major US music companies discontinue sales of pre-recorded Compact Cassettes, and only 17.2 million are sold in the US (down from a peak of 450.1 million in the US in 1988) (2003)

HitClips Disc (2003 – 2004)

Compact Disc sales peak in the UK at 162.4 million units, and decline each year afterwards (2004)

Hi-MD (2004 – 2011)

Sony stops production of Digital Audio Tape (DAT) recorders (2005)

DualDisc (2005 – 2009)

The first music on USB memory sticks is launched in the UK (2006)

USB flash drive (2006 –)

Some albums are sold on microSD memory cards. Some are on generic cards, and later others under brands such as Gruvi, slotMusic and MQS (2007)

microSD card (2007 – late 2000s)

Super High Material CD (2007 –)

Tooth Tunes (2007 –)

VinylDisc (2007 –)

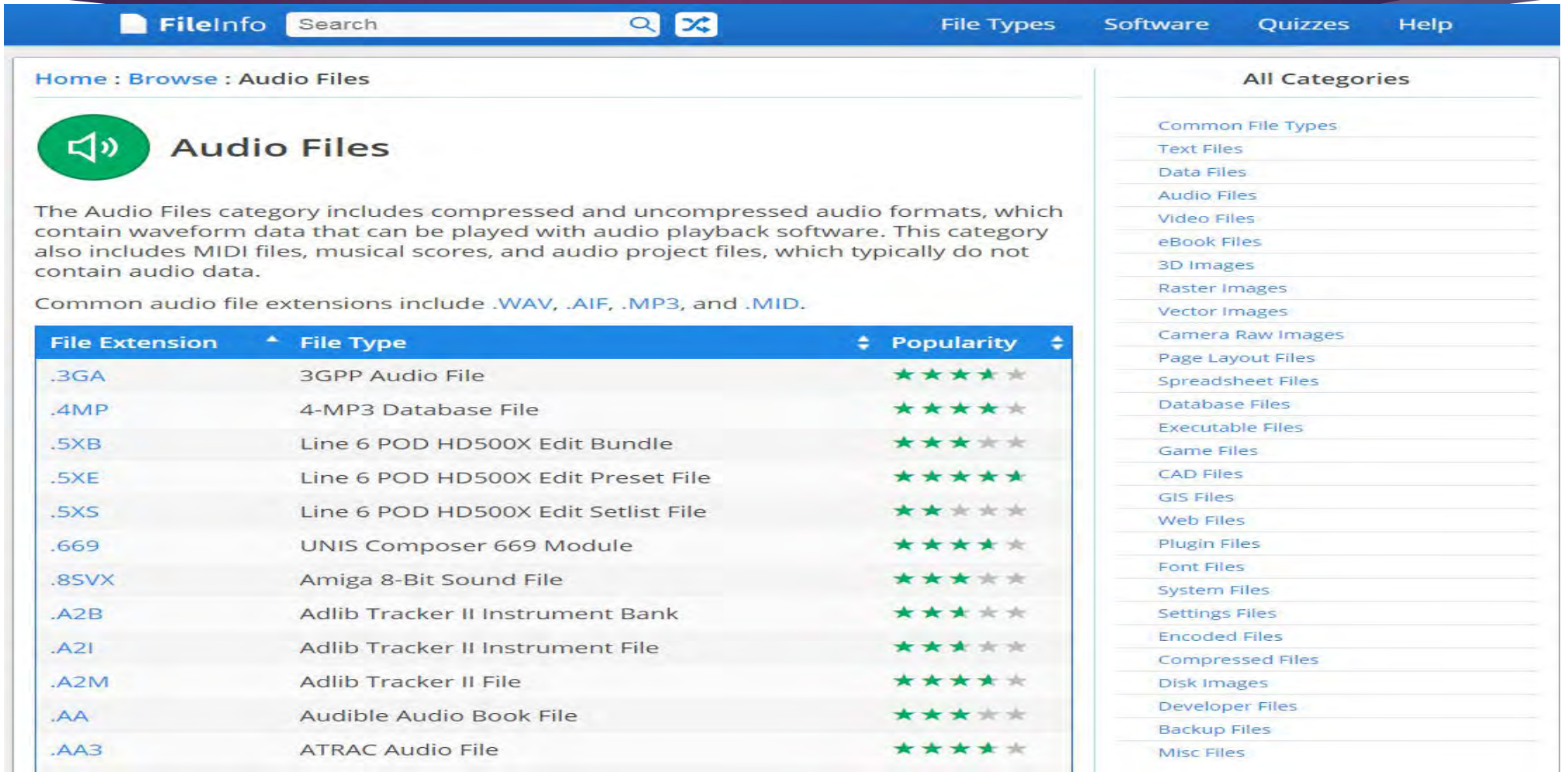
Mass production of piano rolls ends as MIDI files replace them in player pianos (2008)

Spotify is launched (2008)

slotMusic (2008 – 2012)

Blu-spec CD (2008 –)

Audio Formats



FileInfo Search File Types Software Quizzes Help

Home : Browse : Audio Files

Audio Files

The Audio Files category includes compressed and uncompressed audio formats, which contain waveform data that can be played with audio playback software. This category also includes MIDI files, musical scores, and audio project files, which typically do not contain audio data.

Common audio file extensions include .WAV, .AIF, .MP3, and .MID.

File Extension	File Type	Popularity
.3GA	3GPP Audio File	★★★★½☆
.4MP	4-MP3 Database File	★★★★★☆
.5XB	Line 6 POD HD500X Edit Bundle	★★★☆☆☆
.5XE	Line 6 POD HD500X Edit Preset File	★★★★★½
.5XS	Line 6 POD HD500X Edit Setlist File	★★☆☆☆☆
.669	UNIS Composer 669 Module	★★★★½☆
.8SVX	Amiga 8-Bit Sound File	★★★☆☆☆
.A2B	Adlib Tracker II Instrument Bank	★★½☆☆☆
.A2I	Adlib Tracker II Instrument File	★★½☆☆☆
.A2M	Adlib Tracker II File	★★★★½☆
.AA	Audible Audio Book File	★★★☆☆☆
.AA3	ATRAC Audio File	★★★★½☆

All Categories

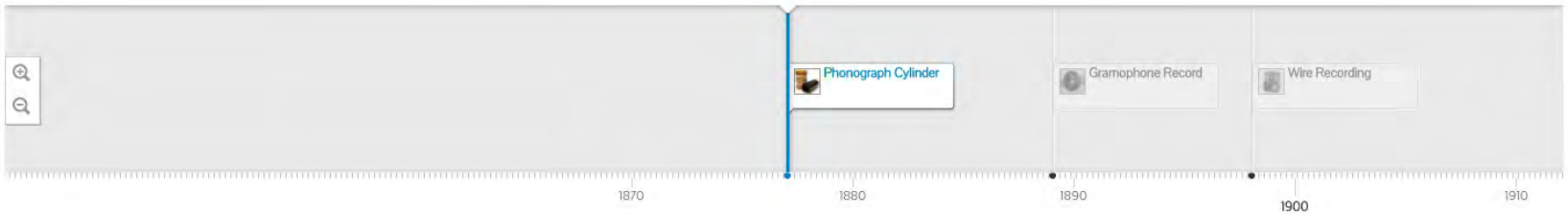
- Common File Types
- Text Files
- Data Files
- Audio Files
- Video Files
- eBook Files
- 3D Images
- Raster Images
- Vector Images
- Camera Raw Images
- Page Layout Files
- Spreadsheet Files
- Database Files
- Executable Files
- Game Files
- CAD Files
- GIS Files
- Web Files
- Plugin Files
- Font Files
- System Files
- Settings Files
- Encoded Files
- Compressed Files
- Disk Images
- Developer Files
- Backup Files
- Misc Files

Audio Formats for Music

1877

Phonograph Cylinder

The grandfather of the physical format, the phonograph cylinder or Edison cylinder was the earliest commercial medium for recording sound. Amusingly, when the disc record was introduced a few years later, many felt the engraved cylinders sounded better. Nothing's changed then.



The evolution of physical music formats - an interactive timeline

Video Formats

Video Format Timeline




A brief history of video recording and playback, from the 1950s onward, including details of all the video formats in the Museum.

Dates of individual formats are approximate and refer to availability in the UK or Europe where known, otherwise for US or elsewhere.

1950s

The first commercially successful video tape format for broadcast use, 2-inch quadruplex (or quad), is introduced by Ampex (1956)

The BBC conducts a live demonstration of the Vision Electronic Recording Apparatus (VERA) videotape recording system it had been developing since 1952, but scraps it soon after in favour of quadruplex (1958)

1960s

Ampex 2-inch helical scan video tape (1961 – 1970)

Sony EV 1-inch open reel video tape (1964 – early 1970s)

Akai ¼-inch open reel video tape (1969 – late 1970s)

EIAJ-1 ½-inch open reel video tape (1969 – early 1980s)

1970s

Sony launches the first video cassette format, U-matic (reaching the UK in 1973). Originally intended for the domestic market, it finds a market in industrial, educational and broadcast applications (1971)

Cartrivision in the US is the first video format to offer films for rental (1972)

Cartrivision (1972 – 1973)

EIAJ-2 (1972 – late 1970s)

V-Cord (1972 – late 1970s)

Video Cassette Recording (VCR) (1972 – 1979)

U-matic (1973 – 1990s)

1990s

Video Single Disc (1990 – 1991)

Laser Juke (1990 – 2002)

Sony ceases production of Betamax VCRs for the US market (1993)

CD-i Digital Video (1993 – 1994)

Digital Betacam (1993 – 2016)

Video CD (1993 – 2000s)

D5 /D5 HD (1994 – late 2000s)

DVD-Video is launched in Japan (1995)

Digital-S / D-9 (1995 – early 2000s)

MiniDV (1995 – late 2000s)

DVCPRO (1995 – early 2010s)

Betacam SX (1996 – 2007)

DVCAM (1996 –)

DVD-Video is launched in the US (1997)

MovieCD (1997 – late 1990s)

HDCAM (1997 – 2016)

MiniDVD-R (1997 –)

DVD-Video is launched in the UK and Europe (1998)

DVD-Video (1998 –)

Interactive DVD (1998 –)

DVD-10 / double-sided DVD (1998 –)

DIVX (Digital Video Express) (1998 – 1999)

Sony Ruvi (1998 – 1999)

D-VHS (1998 – 2007)

Digital 8 (1999 – 2007)

DVD single (1999 – 2010)

2000s

MicroMV (2001 – 2006)

Superbit (2001 – 2007)

MPEG IMX (2001 – 2016)

Worldwide production of DVD-Video discs surpasses that of VHS tapes (2002)

Sony stops producing Betamax video cassette recorders (2002)

D-Theater (2002 – 2004)

Personal Video Disc (PVD) (2003 – 2006)

Flexplay (2003 – 2009)

HDV (2003 – 2011)

HDCAM SR (2003 – 2016)

Professional Disc (2003 –)

Nintendo Game Boy Advance Video (2004 – 2007)

Universal Media Disk (2004 – 2011)

DualDisc (2005 – 2009)

HD DVD and Blu-ray, two competing high-definition optical disc formats are launched, starting a brief format war that ended in early 2008 when Toshiba ceased development of the HD DVD format (2006)

HD DVD (2006 – 2008)

HD DVD/DVD Combo Format (2006 – 2008)

Blu-ray Disc (2006 –)

Blu-ray Disc Recordable (BD-R) (2006 –)

Blu-ray Recordable Erasable (BD-RE) (2006 –)

Netflix launches its streaming video service (2007)

JVC, the company that invented the VHS format, ceases production of standalone VHS video cassette recorders (VCRs) (2008)

Toshiba announces it will no longer manufacture or market HD DVD players or disc drives, ending the format war with Blu-ray (2008)

Pioneer ceases production of its remaining LaserDisc players (2009)

Film Formats

Film Formats

Film formats for still photography and motion picture recording and playback.

Dates are approximate and refer to availability in the UK or Europe where known, otherwise for US or elsewhere.

17th Century

Magic lantern (17th century – 1940s)

1830s

The daguerreotype process, using silver-plated copper sheets treated with iodine, becomes the first publicly available photographic process (1839)

William Henry Fox Talbot introduces the calotype or talbotype photographic process, using paper coated with silver iodine (1841)

1850s

The collodion process is invented, replacing the daguerreotype process by the end of the 1860s (1851)

Ambrotype (early 1850s – 1880s)

Tintype (1850s – 1930s)

Photographic plate (1851 – 1990s)

Carte de visite (1854 – 1900s)

Stereoview (1850s – 1920s)

1860s

Cabinet card (1863 -early 1920s)

1870s

Cigarette card (1875 – early 2000s)

1880s

George Eastman introduces the first transparent plastic roll film, using a nitrocellulose base (1889)

1920s

9.5mm film (1922 – 1960)

16mm film (1923 –)

1930s

620 film (1932 – 1966)

135 film (1934 –)

View-Master (1939 –)

1940s

Filmstrip (1940s – 1980s)

The first Polaroid instant film camera, the Model 95, is launched (1948)

1950s

Slide carousel (1950s – 2000s)

Lestrade (1954 – 1970s)

Vistascreen (1955 – 1960s)

Give a Show (1959 – early 1980s)

1960s

Pathescope Ltd. goes into liquidation, ending production of 9.5mm film in England (1960)

Microfiche (1961 – 2000s)

126 film (1963 – 1988)

Show'N Tell (1964 – 1980s)

Americom 8mm Home Movies (1965 – late-1960s)

Single-8 (1965 – 2012)

Super 8 (1965 –)

Technicolor Sound Movie Cartridge (1968 – 1970s)

Tape-slide set (1960s – 1980s)

La Belle Commpak (late 1960s – early 1980s)

1970s

Kalavox sound slide (1970s)

Talking View-Master (1970 – 1981)

Movie Viewer (1971 – 1985, 2014 -)

Polaroid SX-70 film (1972 – early 1980s)

110 film (1972 – 1990s)

Polaroid sues Kodak for copyright infringement with their instant cameras (1976)

Polaroid Polavision (1977 – 1979)

1980s

Polaroid 600 film (1981 – 2008)

Disc film (1982 – 1988)

Talking View-Master Electronic 3-D Viewer (1984 – late 1980s)

1990s

The JPEG standard for digital images is finalised (1992)

Photo CD (1992 – 2004)

Advanced Photo System (APS) (1996 – 2004)

Talking View-Master 3-D (1997 – 1998)

Fujifilm Instax Mini (1998 –)

Polaroid i-Zone (1999 – 2006)

2000s

Kodak ceases production of cameras using APS (Advanced Photo System) cartridge film (2004)

Kodak makes its last carousel slide projector (2004)

Polaroid stops producing film for its instant cameras (the Dutch company The Impossible Project is subsequently formed to continue manufacture of some instant films) (2008)

2010s

Paramount's last movie on 35mm film is released (2013)

View-Master Virtual Reality Experience Pack (2015 –)

Digital Data & Digital Media

Digital Data is stored on three types of media

- **Magnetic**: Magnetic disks include the hard drive on your laptop, external hard drives, network environments and servers, and magnetic tape (reel-to-reel and cartridge).
- **Optical**: Optical media include Compact Disks (CD, CD-ROM, CD-R, CD-RW), Digital Versatile Disks (DVD, DVD+R, DVD-R, DVD-RAM, DVD+RW, DVD-RW), Write-Once, Read-Many (WORM) disks, High-definition DVD (Blu-ray & HD-DVD), Smart Cards, and Optical tape.
- **Solid State**: Flash memory cards, USB Flash drives, and Solid State Hard Drives.

Digital Media Lifespan

What is the lifespan of different types of digital media? It depends on many factors:

- Original Quality
- Storage conditions: temperature, humidity, air, moisture, light.
- Age and Handling
- Frequency of access

Manufacturers will tell you that a HDD will last 30 years. Real world usage shows only 3-5 years. Flash storage should last 5-10 years, but high usage (read-write cycles) will shorten that considerably. CD's and DVD's have a shelf life of 5-10 years unrecorded, but only 2-5 years recorded. Manufacturers said they would last 100 years! Magnetic tape will last 10-20 years under ideal conditions (stable temperature and humidity).

The other part of the problem: media obsolescence. Can you access the older media with current technology? Do you have access to the older technology?

Best Practices: Data Storage

Data Storage can seem pretty simple. Just store the data on your hard drive or in the Cloud, right? You should ask yourself some questions about the data before you start making decisions about storage.

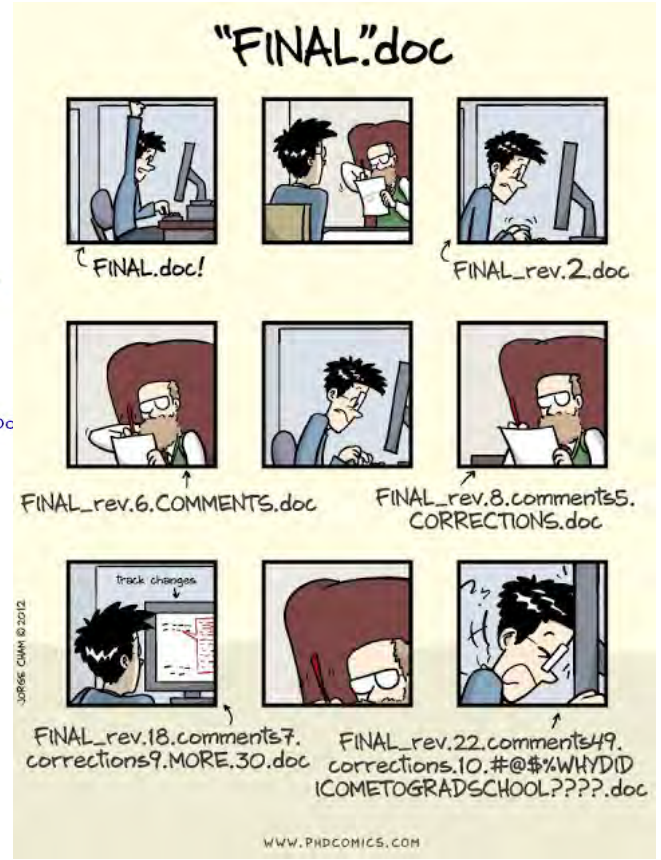
- How important is the data?
- Do I need to keep this data? Can the data be reproduced, or is it unique?
- How long do I want or need to keep the data?
- How fast do I need to access the data?
- How secure do I need to keep the data?
- Do other people need to access the data?
- What institutional or funder requirements need to be adhered to?

Best Practices: First Steps

- Locate all of your data files that you want to store.
- Decide what you need to keep.
- Create a directory identifying your data files by name, format, size, and file location. Keep it current!
- Label the data containers, and the media in the containers. What happens if they get separated?
- Locate and/or create supplemental documentation (metadata) for each data file. Include variable names, descriptions, units, standards, instrument calibrations, codes, algorithms used to transform the data, software (including version and OS). What do you need to be able to use this data file if you do not remember?
- Organize the files. Document your organization methodology. Use it consistently.

Best Practices: File Naming

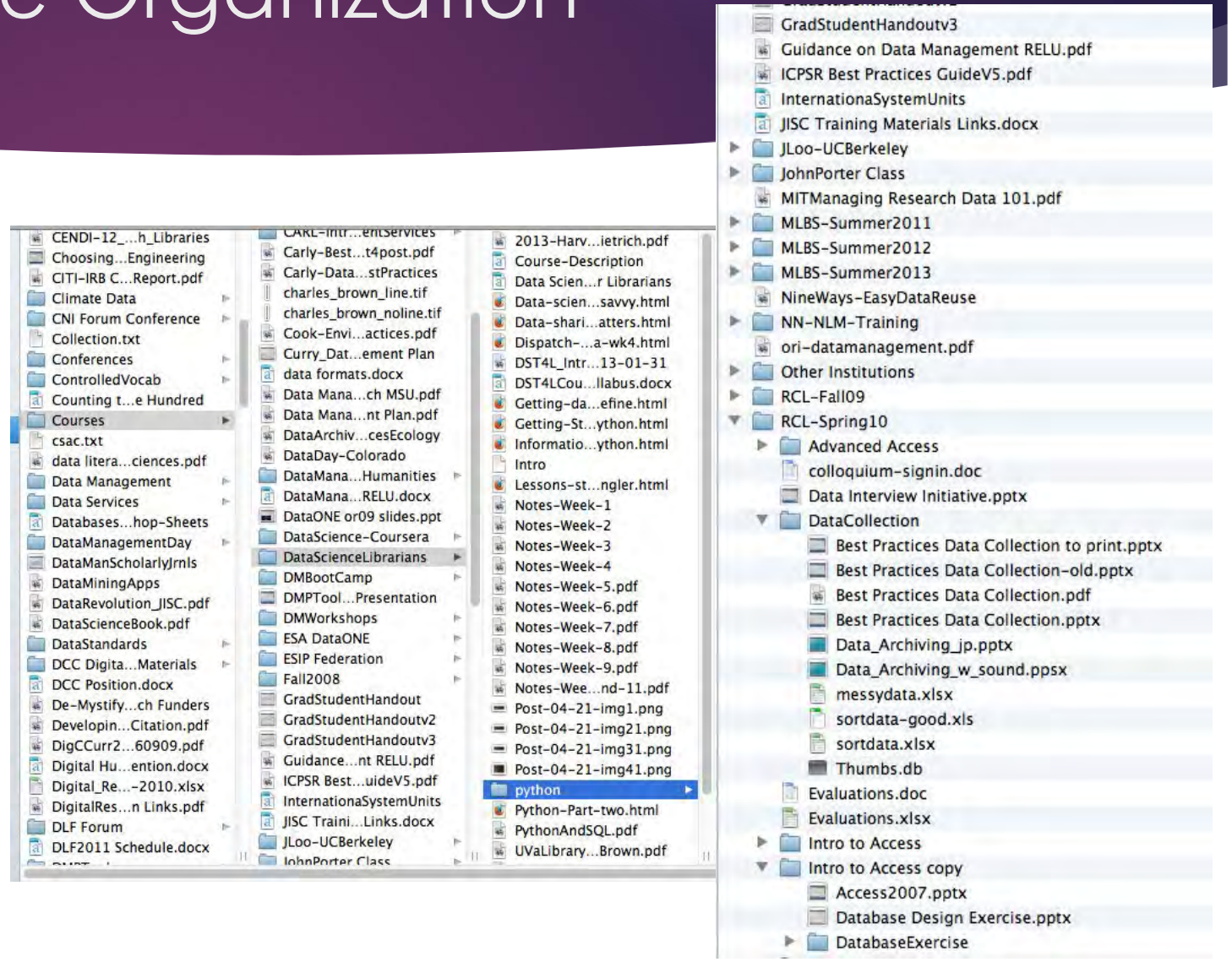
- **Be Consistent**
- Use descriptive names
- Not too long (32 characters max); CamelCase
- Try to include time
- Date using YYYYMMDD (create chronological order)
- Use version numbers
- Don’t use special characters [*&^$!~]
- Don’t use spaces - use “-” or “_”
- Don’t change default extensions
- Identify different versions clearly
- Add zeroes in front for large data sets (000001 instead of 01 if you expect 10000 images)



http://www.phdcomics.com/comics/archive.php?comicid=1531

Best Practices: File Organization

- Folders named for major functions/activities
- Structure by date or event (especially subfolders)
- Names should be self-explanatory
- Avoid duplication
- Make it simple & consistent



Best Practices: Primary Data

Never work with your primary data file! Always make a copy to work with. Your computer hard drive or working environment should only store your current working data file. Your primary (master or raw) data file should be stored in a safe environment, and backed up.

Disasters and accidents do happen:

Hardware failures	Software problems
Virus Infections	Corrupted data files
Power failures	Hacking
Stolen computers	Human error
Natural Disasters	Media degradation

Keep the original file as an read-only file. Give it a file name that can be used as the first part of all subsequent files related to it.

Best Practices: 3-2-1-Rule & Threat Zones

Data Backups are very important. Follow the 3-2-1 Rule:

- Keep three copies of any important data files – a primary and two backups.
- Keep two copies on different digital media – a HDD and Flash drive.
- Keep one copy offsite, or at least offline.

This is sometimes called Here-Near-Far. Working copy Here, primary backup Near, and second backup Far.

What is a **threat zone**? A different geographical location from the one you are working in. If a natural disaster occurred in Virginia, and all of your data files are here, then you will probably lose everything. Put your primary, or primary backup, on a media that is secure, and send it to a friend or colleague in another state for safe keeping. Can the Cloud be considered a different zone? Yes, if the servers aren't in the same zone you are in. Some cloud storage providers allow you to specify where your data is physically stored.

Best Practices: Backups

UVa has a license for CrashPlan, a cloud-enhanced desktop backup service from Code42. It is available to staff, faculty, instructors, and degree-seeking graduate students. It is only to be used with public and moderately sensitive data. The CrashPlan FAQs provide information including getting started and backup-related sections. CrashPlan has an extensive knowledge base available with guides, troubleshooting and a configuring help. It is available for both Windows and Mac machines.

UVa had made CrashPlan available to home users with a 25% off deal with Code42. However, as of 10/23, CrashPlan Home is no longer available. They recommend Carbonite or CrashPlan for Small Business.

ITS also provides backup for the servers it manages.

There are many providers of Backup services available, both online and software-based. Automatic backups are a great option for ensuring that your data is always protected. Acronis, Paragon, and StorageCraft are some of the better companies in this space.

Best Practices: Backup Validation

Manually check 5-10% of data files yearly.

- Is the file collection complete? Compare it to your directory of data files.
- Did the files transfer properly? Do a bit-by-bit comparison of random files.
- Use a MD-5 checksum (hash). The UK Data Archive has a good exercise that will help you understand checksum.
- Write-once media validation. Did you create a validation hash when you created optical disks? If so, compare.
- Volume and directory validation. Check the media for directory or volume corruption.
- Storage media integrity. Run a media scan on your HDD for bad sectors.
- Visual inspection. Does the media look ok?

Best Practices: Network Security & Access Control

- **Network security:** Keep confidential or highly sensitive data off computers or servers connected to the internet
- **Physical security:** Access to buildings and rooms
- **Encryption:** Provides protection by scrambling data, so only the owner of the key or password can read the data. This protects the confidentiality of the data so that if an unauthorized person gained access to the storage device or service, they would be unable to see the data. It also protects the integrity of the data so that it cannot be tampered with without the owner knowing it.
- **Computer systems & files:** Use strong passwords on files and systems; use Virus protection (updated continuously and running!); Do not send personal or confidential data via email or FTP. Transmit as encrypted data and require data access agreements or confidentiality agreements from recipients.
- **VPN**: A VPN scrambles data as it is transmitted between your mobile device and a server. This allows you to access sensitive data securely stored on a remote server. UVa offers several levels of VPNs. Some are designed to handle secure data.

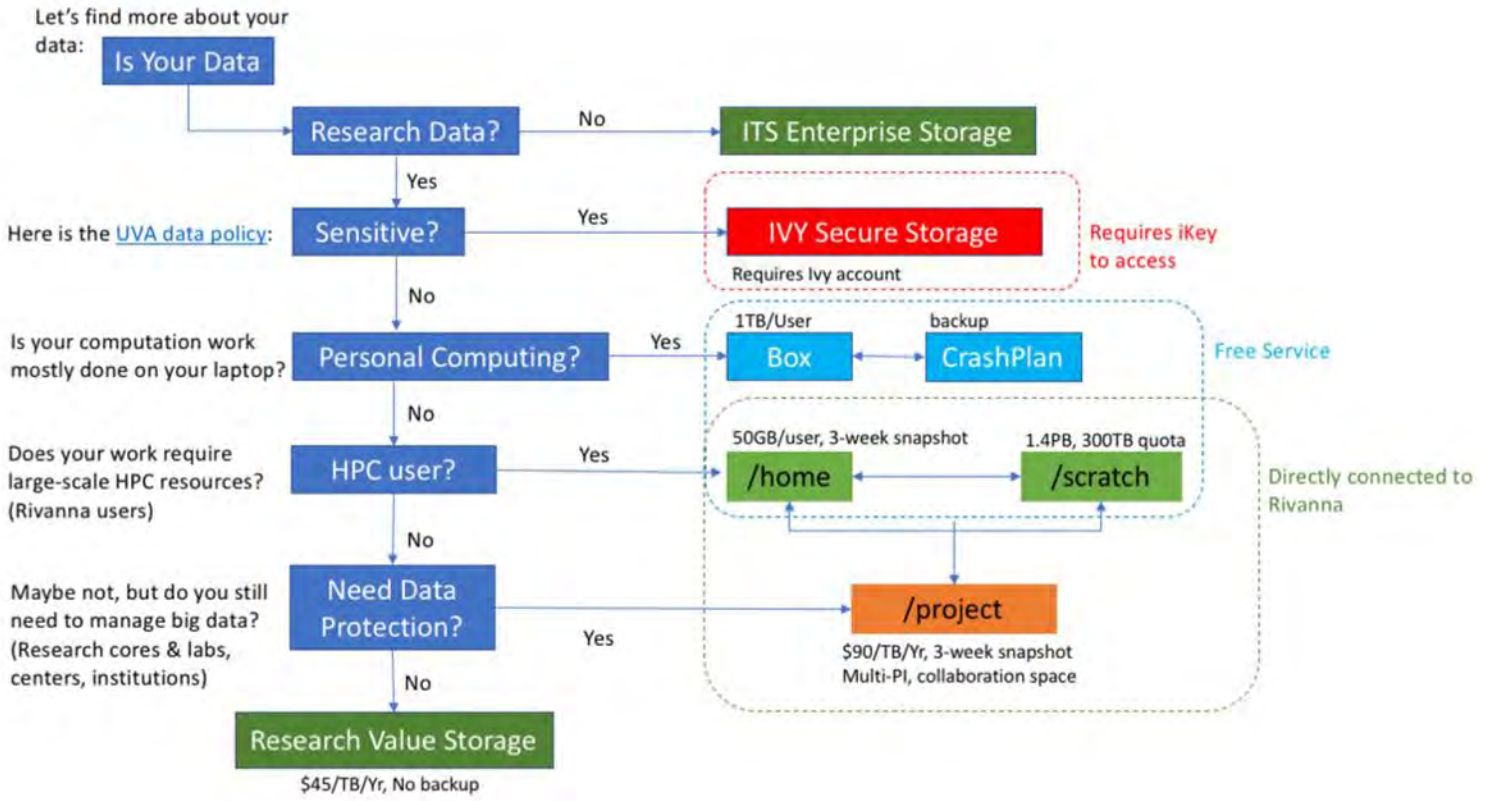
Best Practices: Common Sense Tips

- Desktops, tablets, laptops, and phones should NOT be used for storage of your raw, original, or only copy of data.
- Removable media – USB Sticks, Flash Cards, Memory Cards, CD's, DVD's, Cassettes, DAT's, portable external HDD's - should NOT be used for storage of your raw, original, or only copies of data.
- All removable media are subject to degradation and failure. It **will** happen. Removable media are all inherently vulnerable to temperature and humidity fluctuations, poor handling, air, moisture, light conditions, theft, mechanical breakdowns, forgetfulness.
- Manage your stored data. Visit often, at least once a year. Migrate your data media to new media on a pre-set schedule. Migrate to newer formats when possible. Migrate to newer software if possible. Always verify data consistency.
- Keep your directory up to date. Keep a current copy with your data storage options.

UVa Data Storage Options - Research

Research Data Storage

Visit the CADRE site for information about the Various options available At UVa.
They manage the Rivanna and HPC resources.

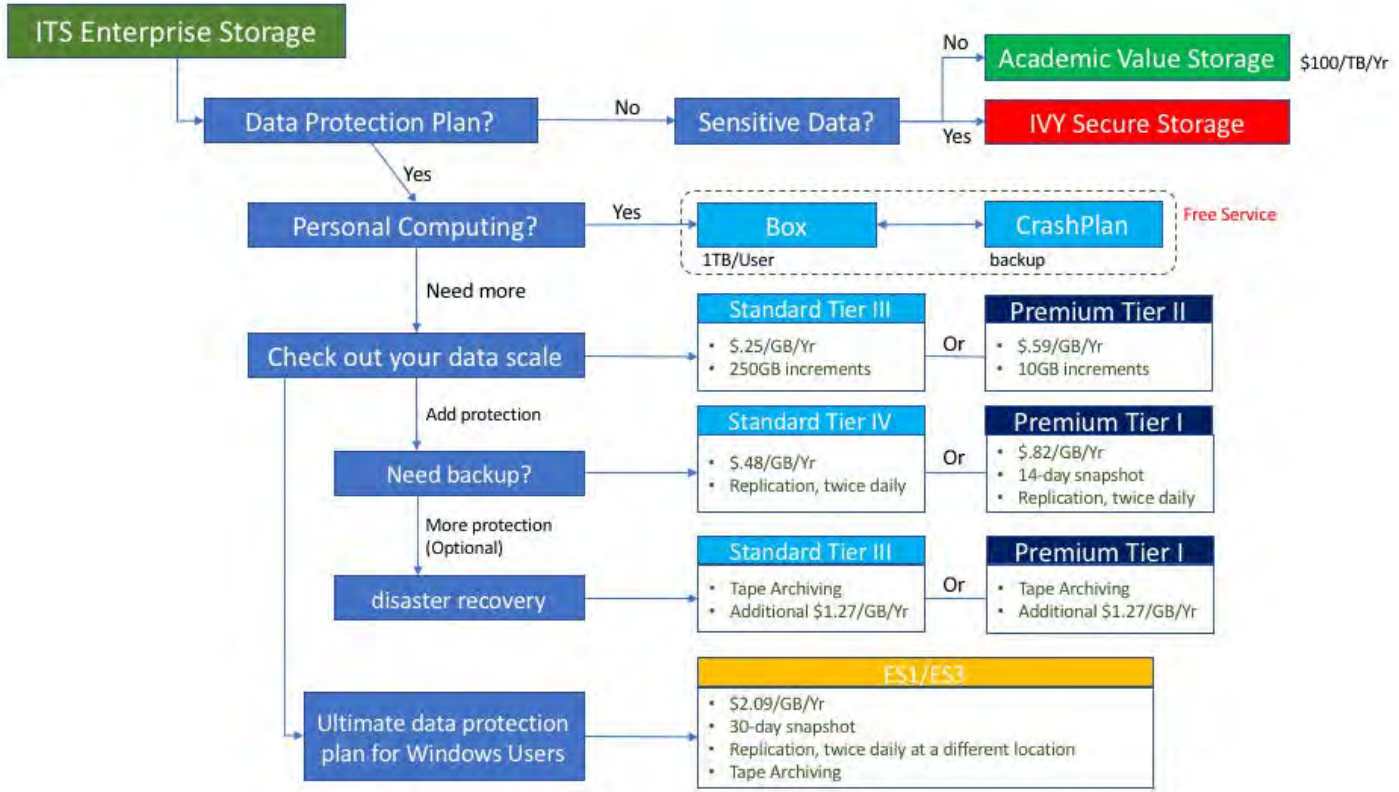


https://cadre.virginia.edu/service-detail/storage

UVa Data Storage Options - Enterprise

Enterprise Data Storage

Visit the CADRE site for information about the Various options available At UVa.
They manage the Rivanna and HPC resources.



https://cadre.virginia.edu/service-detail/storage

UVa Data Storage Options - ITS

UVa's ITS provides three levels of storage. All are fee-based:

- Standard: Standard Storage with No Data Recovery (Tier IV), Standard Storage with Data Recovery (Tier III), and Standard Storage with Data Recovery (Tier III) and Backups.
- Premium: NAS High-performance storage. Premium Storage with No Data Recovery (Tier III), Premium Storage with Data Recovery (Tier I), Premium Storage with Data Recovery (Tier I) & Backups, Windows Environment Premium Storage (ES1, ES3)
- Value: Self managed for sensitive & non-sensitive UVa data. Academic Value Storage, and Research Value NAS Storage (researchers only).

They provide RESSCU service to store mission critical electronic data at a remote storage facility in Blacksburg, VA. This is a co-owned (with Virginia Tech) Hierarchical Storage Manager (HSM)

UVa Data Storage Options - Personal

- Personal computer
- UVa Box 1TB storage
- UVaCollab While not intended for data storage it provides 4 GB storage. Secure, behind NetBadge.
- Home Directory Service 4 GB storage
- External Hard Drives (HDD), Flash, USB drives
- Amazon Web Services
- OneDrive
- SharePoint Online
- Dropbox
- Google Drive

UVa Data Storage Options – UVa Box

UVa Box is a free cloud-based storage and collaboration service available to all UVa eligible students, faculty, and staff. You can store up to 1 TB of non to moderately-sensitive data in it. It is accessible from anywhere you have internet access. You can integrate applications into Box.

UVa Box is the same as Box – UVa has an signed agreement about data (which is why the medical folks don't have access to it). But a non-UVa Box user can interact with a UVa Box user.

The UVa Box FAQs can answer many of your questions.

How to use UVa Box includes instructions for working with files, sharing & collaborating, and some of the Box features.

The UVa Box User Responsibilities & Data Restrictions includes the information you need to understand what types of data you can put in your UVa Box account.

The Box Community is a great place to learn about Box from the Knowledge Base and other Users.

UVa Data Storage Options – UVa Collab

UVa Collab is the LMS – Learning Management System – at UVa. It is the University of Virginia's central online environment for teaching, learning, collaboration, and research. UVACollab partners with faculty, staff, and students in the work that sustains the Academical Village—engaging in interactive discussions, joining virtual meetings, securely storing and sharing materials, and much more.

Each site has a storage capacity of 4 GB. You can have multiple sites. It is secure (behind NetBadge) and you can collaborate with colleagues at non-UVa institutions.

It is not designed to be used as a storage option, but 4 GB is sufficient for a lot of folks. Great for smaller projects! You can link to external sites and resources, and it has several integrated tools including 3rd party ones - a WordPress Blog and Confluence Wiki. The Kaltura Media Gallery is a good too if you work with images and videos.

UVaCollab has a extensive Knowledge Base available when you need assistance.

UVa Data Storage Options – Home Directory

UVa Home Directory provides:

- Convenient online file storage - 4 Gigabytes (GB).
- Backup copies of your documents - within certain time limits.
- Easy Web publishing - publish and manage webpages. Your personal webpage URL is http://people.virginia.edu/~Your UVA Computing ID.
- Easy access using a software program on your computer or via the Internet from any computer connected to the UVA network.

There are FAQs for Windows users and Mac users. Files stored in Home Directory are backed up daily. File Recovery information can help you recover from a loss. Mapping a drive in Windows or Mac OS X is a useful feature that allows you to direct-connect to transfer files to and from your desktop.

UVa Data Storage Options – SharePoint Online

SharePoint Online is part of the Office 365 suite of tools from Microsoft. ITS also offers Groups & Teams which include a limited version of SharePoint. The Jumpstart Guide for Microsoft Teams is a good place to start if you are interested. ITS also has information about the differences between the two products.

The Microsoft SharePoint Online website provides a lot of information about this tool.

SharePoint Online is great for:

- Content management
- Customization of HTML/CSS to change look and feel
- Strong use of permissions and inheritance in multi-site layout
- A site template other than the default (Publishing site, Communication Site).

SharePoint Online is accessible from any internet connection, and the permissions allow great granularity to protect your data. It is integrated with the other Office 365 applications.

UVa Data Storage Options – External Drives

There are many types of external hard drives available. Remember that if you are storing UVa-related data (funded project, UVa data, etc.) on one you need to follow the UVa Data Protection Standards (3.0): https://security.virginia.edu/university-data-protection-standards .

Best Practices: Cloud Storage at UVa

Cloud storage can be used to store additional copies of data. Be very careful reading and understanding the terms of use, privacy policies, data ownership policy, and fee structure.

Cloud storage is not appropriate for highly sensitive data (HIPPA). Moderately-sensitive data (FERPA) may be allowed if the provider has an agreement with the University. Be sure you are adhering to the University Data Protection Standards (UDPS 3.0) at https://security.virginia.edu/university-data-protection-standards if you are storing funded research data or UVa data.

The Information Security office at UVa should be contacted if you have any questions about storing your data. https://security.virginia.edu/

UVa Data Storage Options – AWS

Amazon Web Services provides a suite of tools to do everything with your data. UVa has a signed agreement with DLT Solutions, a 3rd party provider who is a Premier Consulting and Managed Services Partner for AWS. You will need an AWS account set up through ITS to access their services. You will then be able to access the DLT Portal.

DLT has conducted several training sessions here in the last 2 years, and I have copies of the content I can share by request.

AWS provides a long-term storage option – Glacier – which is inexpensive, and would be a good option for data you want to park.

There is a local AWS User Group at UVA. This group is dedicated to connecting UVA students, alumni, and the residents of Charlottesville. Founded and organized by UVA alumni and students, they explore all aspects of working with AWS. Learn about new services and features, hear from developers who are using services in new and exciting ways, learn how to build an engineering career using AWS, and enjoy the company of others who are eager to share experiences.

UVa Data Storage Options – OneDrive

OneDrive is the Office 365 cloud storage service that UVa has an agreement with Microsoft for UVa staff and faculty who use Office 365. Every user has 5 TB of storage available. It is integrated with Microsoft Office Online and many of the desktop applications. It is intended for University-related activities. You can get a personal account if you need one (or three).

Users have access to assistance and training on the UVa Office 365 Resource Portal.

There is an extensive list of FAQs and a page on File Sharing: Internal vs. External.

The User Responsibilities & Data Restrictions page will provide you with the information you need to know about which data files you can store on OneDrive.

	Data *Permitted* in OneDrive	Data *Prohibited* in OneDrive
OneDrive	• University files containing non-sensitive data • University files containing moderately sensitive data, including FERPA-protected data not covered under the definition of highly sensitive data	• Health information, including HIPAA-protected data • Social Security, passport, or financial account numbers • Credit card processing (PCI) data • Export-controlled research (e.g. ITAR, EAR) • Any other highly sensitive data *(refer to University policy for examples)*

UVa Data Storage Options – Dropbox

Dropbox is not an "approved" cloud storage provider for UVa (we do not have a signed agreement with them). But a lot of folks use it. It is similar to Box in how it works and the functionality. There is a free version – Basic, and paid plans in Dropbox Business.

They have an extensive knowledge base Help Center to answer your questions. It includes a Getting Started section.

The Dropbox Community is a great resource.

Remember to read the Terms of Use and Privacy policy for any Cloud-based service provider, even if they are approved by UVa.

UVa Data Storage Options – Google Drive

Google Drive is part of the Google Gmail package that UVa has for students and alumni. It is not recommended for staff and faculty use, though if you interact with the students on a regular basis it makes sense to have an account.

The Gmail FAQs provide a lot of information, and Google Gmail Help is also available. The Google apps that UVa supports with Gmail are: Contacts, Drive and Docs, Calendar, Sites, Talk, and Video.

Drive provides 15 GB of storage on personal accounts.

Remember to read the Terms of Use and Privacy policy for any Cloud-based service provider, even if they are approved by UVa.

UVa Data Storage Options – Azure

UVa does not currently have an agreement with Microsoft for the Azure Computing Platform. Azure is a set of cloud services similar to the AWS system.

Research Data Services + Sciences

I am part of the Research Data Services + Sciences unit of the UVa Library. Our website has information on all of our services.

The Research Data Management section includes a link to my Research Data Management Subject Guide.

We offer an assortment of workshops each semester. We also have an archive of downloadable content from previous workshops.

We are part of the CADRE group.

Thanks for attending!

Do you have additional questions? Feel free to contact me at wtc2h@Virginia.edu to set up a consultation.