

# Computer Aided Archaeology

## 02 - Your Computer as research tool

Martin Hinz

Institut für Archäologische Wissenschaften, Universität Bern

06/10/21

# Software suggestions

- Office: [Libre Office](#)
- Image Processing
  - Raster: [Gimp](#)
  - Vector: [Inkscape](#)
  - Digital Photos: [Darktable](#)
- Layout: [Scribus](#)
- Citation Manager: [Zotero](#)
- GIS: [Quantum GIS \(QGIS\)](#)
- Statistics
  - GUI: [PAST](#)
  - script: [R](#) & [RStudio](#)
- 14C Calibration: [OxCal](#) (or [Chronomodel](#))
- digitizing charts: [Engauge](#)







# Online Ressources

- Translation: [DeepL](#)
- file transfer [Swisstransfer](#)
- Polls ('doodle'): [dfn terminplaner](#)
- literature search:
  - Switzerland: [swisscovery](#)
  - worldwide: [WorldCat](#)
  - ancient studies: [Zenon](#)
- (fast) online calibration: [CalPal online](#) (currently not working)
- 14C Dates: [XRONOS](#)
- archaeological data: [Journal of Open Archaeology Data](#)

## limited recommendability

- cloud hosting [Dropbox](#)
- Online Office [Google Docs](#)

# Working with Digital Data

-  File structures - *where to put stuff so you won't lose it*
-  File naming - *what to call stuff so you know what it is*
-  Version control - *keeping track of stuff*
-  File formats - *what to save stuff in so it's safe* (Text, Images, Spreadsheets/databases, CAD/GIS, Audio/Video)
-  Documenting data - *letting others understand your data*
-  Selection - *chucking stuff away!*

# 📁 File Structure

Where to put stuff so you won't lose it

- Logical to you – and easily understandable to others
- Ease of sharing / exchange of data
- Maintaining retrieval of files, e.g. Esri shape files (multiple files)
- Defining the 'end product' of a project helps maintain file structure

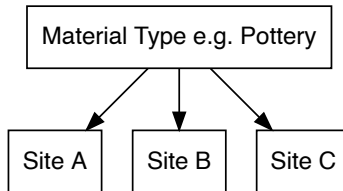


Name
analysis.R
data-cleaning.R
protocols.pdf
raw_data.csv
variable_guide.pdf

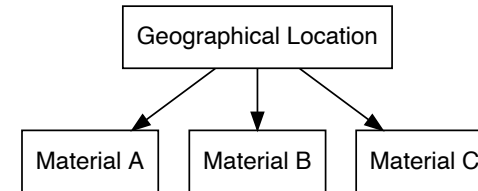
Name
Final
Old code
Analysis code.R
Analysis code w revisions 3.7.18.R
Data_april.csv
Data_april_BAB.csv
Data_april_final.csv
Data_april_final (copy).csv
Data_may.csv
regressions.R

# Which primary data defines your research?

## Archaeological material



## Location (site based)

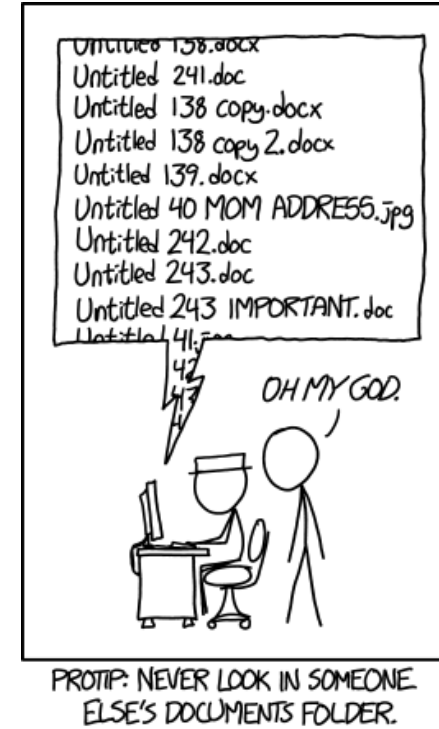


- Distinguish between projects.
- Distinguish between sub-folders.
- Define 'end-product' of research – and keep clean of temporary folder and files.
- Research designs change and so must file structure.
- Avoid overuse of folders – easier said than done though.
- rule of thumb: 7 subfolders, if more, structure deeper

# File Naming

What to call stuff so you know what it is

- Names tell us what a file is: Contextual information.
- Names order files: Making stuff easy to find.
- Define your system: **And stick to it.**



source:<https://xkcd.com/1459/>



# File Naming (suggestions)

First: Define the types of data and file formats for the research.

- Different data may require different naming conventions:
  - Should different data/file formats be identified as part of same project?
- Examples of contextual information in file names:
  - Date, Author or Initials, Site or Project, Material.
- Capitals in file names affect ordering – be consistent.
- Numbers order files only if zeros are used before units and tens:
  - 001, 002, 003, etc will order files up to 999.
- Dates are useful for version control and ordering files.
  - YY-MM-DD (11-03-02) at end of name orders files of same name by year.
  - Year first is good for ordering files, e.g. publication pdfs
- Spaces between file names cause havoc in GIS. Use underscores
- / Forwardslashes / in file names can cause problems too.
- CAPITALS ARE HARD TO READ!



# Version Control

## Keeping track of stuff

It's surprisingly easy to lose track of the current version of a file.

Especially:

- Word file drafts of thesis chapters.
- Word files commented on by others.
- Multiple-author files sent back and forth by e-mail.
- Graphics and AutoCAD files.
- Be consistent with up-dating file names: version number, initials, date.
- Put old versions in separate "Archive" folder (with date of that version in the folder name)
- Possibly delete old drafts when final version is finished.

# File Formats

What to save stuff in so it's safe

- Facilitate exchange of data
- Ease of working on different computers / software packages
- Preserve data for re-use in the future



source: <https://roaringelephant.org>

# File Formats

## Key Issues

Closed	vs	Open
Proprietary	vs	Open
Non-Standard	vs	ISO Standard
non-human readable (binary)	vs	human readable (text based)
Compressed	vs	Uncompressed

# Text Files

- Manuscripts produced on computers: word files.
  - conference notes, articles, theses, books, etc.
- Scanned printed material often made into a PDF file.
  - Conversion into editable text files using Optical Character Recognition (OCR) software.
- Marked-up formats for viewing as web-pages: HTML
- Other Mark-up languages
  - LaTeX
  - Markdown

# Text Files

Format	Description / Properties	Usage and Archival Recommendations
.txt	Text file. Simple plain text document. Compatible across software packages. Supports very little formatting.	Good for extremely simple files. Commonly used for introductory “Read me” files containing basic information on project archives.
.doc	Microsoft Word document ( - 2003) Proprietary binary format. Can be read by OpenOffice. Easily converted into PDF format.	Accepted for archiving because it is so widely used. However, will soon become obsolete.
.docx	Microsoft Word document (2007) Human readable XML format. Stored along with embedded content as zipped file.	Good for dissemination and preservation. Conversion to .doc file to open with earlier versions of MS Word.
.rtf	Rich Text Format (Microsoft) Tagged plain text format.	Formatting issues when using opened in different software. Large file sizes mean that .docx or .odt file formats are preferred.
.odt	Open Document Text (OpenOffice) ISO standard, human readable XML format.	Open source format good for use, dissemination and archiving. Archive files in uncompressed form. Can open .doc files. Might not open correctly in other word processing programs.
.pdf	Portable Document Format (Adobe) Proprietary binary format. Aims to retain document formatting. Can store embedded data: raster and vector images (e.g. Adobe Illustrator files)	Highly suitable for dissemination. PDF creators and readers freely and widely available. Retain original text document and embedded objects. (e.g. images, tabular data, etc).
PDF/A	Portable Document Format / Archive (Adobe) Open ISO standard format for long term archiving. Formatting data self-contained in file.	Widely accepted as viable format for long-term archiving. Retain original text file and embedded objects separately. (e.g. images, tabular data, etc).

# Digital Images

- Images are used to convey information and support interpretations.
- Images contain data and are often analysed to reach interpretations.
- Images are one of the most important formats of digital data in archaeology.
- Image documentation and preservation is key for future re-use of project archives.
- Digital images often form the largest part of digital archives.
- **Raster Images:** matrix of dots/pixels containing information (photographs, scans)
- **Vector Images:** formed by points, lines, polylines, polygons represented by co-ordinates and mathematical formulae (graphic - illustrations, CAD, GIS)

# Routes to Rasters

- Scanned images of paper illustration or photograph.
- Digitally captured or created: cameras or digitally created illustrations
- Output product of other digital applications: vector, CAD, or GIS work, or geophysical survey data, etc.
- Think of the purpose of image when making it:
  - screen, print or reference image.
- Formats have different qualities which affect their output use and preservation.

# Routes to Rasters

**Resolution** / Level of detail in image:

- Pixels per inch (PPI) or Dots per inch (DPI) or Samples per inch (SPI)
- Bigger the physical size of the picture + increased resolution = bigger file size
- min. 600 dpi for photographs and 300 for illustrations.

**Bit (Colour) Depth** / Level of colour information:

- 2 Bit = Black/White (line drawings with only black and white needed)
- 8 Bit = Greyscale
- 24 Bit = Standard colour
- 32 Bit = High colour

**Colour Space** / Type of colour

- Bi-tonal = black/white
- Greyscale
- RGB (Red/Green/Blue) used for screen presentations. Cameras generally capture images in RGB.
- CMYK (Cyan/ Magenta/Yellow /Key [Black]) used for printing colour images.

**Compression**

- Non-compressed (Lossless): GIF, PNG, TIFF.
- Compressed: JPG

Some formats (TIFF, PNG) allow files to be saved as non-compressed. Important to be aware of when compression is occurring and at what level.

**Image layering** Layering is NOT supported in final raster image and layers will be merged into a single layer from top down.



# Raster File Formats

Format	Description / Properties	Recommendations
.bmp	Bit-Mapped Graphics Format Microsoft proprietary format in older MS applications for simple graphics. Limited embedding of metadata.	Not recommended for either working files or long-term file storage.
.gif	Graphics Interchange Format Compuserve proprietary format. Lossless compression with 8-bit colour depth (256 colours). Limited metadata.	Superseded by PNG format, but still widely used for still and animated Web graphics.
.png	Portable Network Graphics. ISO standard. Lossless compression with 32-bit colour depth, and Alpha channel (transparency), with few 'visible artefacts' (cf. jpegs). Does not support EXIF metadata.	Designed for internet and uses RGB colour space. Standard format for lossless presentation. Use instead of GIF format. NOT recommended long term storage (use TIFF). NOT recommended for digital photographs, as only supports RGB colour.
.jpg / .jpeg	Joint Photographic Expert Group ISO standard. 32-bit colour depth with extremely efficient lossy compression of image. Compression creates 'visible artefacts' around complex high-contrast image areas (e.g. text). Supports EXIF and IPTC metadata.	Designed for photographic or painted images with smooth varying tones that do not have sharp contrast. Much smaller file sizes than PNG or TIFF. While unsuitable for long-term storage, accepted format for archiving digital photographs.
.tif / .tiff	Tagged Image File Format (Adobe) Uncompressed image format. Can support range of metadata: EXIF, GeoTIFF for georeferencing.	Uncompressed Baseline TIFF Version 6 standard format for long term preservation of digital figures.
.psd	Photoshop Document (Adobe) Proprietary format and can be used with open Photoshop Elements software. Supports variety of features: image layering, transparency, text. Supports IPTC, EXIF and XMP metadata.	'Industry standard' for image creation. Proprietary nature means limited third-party support for PSD format. Limited compression results in large file size. Unsuitable for long term preservation. (TIFF for figures or DNG for photographic images.)
raw	Unprocessed bitmap files created by digital cameras and some scanners. Proprietary, require specific software. No standardisation in file formats.	If possible, convert raw files to DNG format for long-term preservation.

# Raster Files

## key points

- Think of purpose of image.
- Document rationale of image creation.
- Maintain image documentation:
- Image properties, file naming and image description file.
- If working with JPEGs, save original as archive and work on copies.
- Save working copies of raster outputs of vector files and replace with final version.

# Vector File Formats

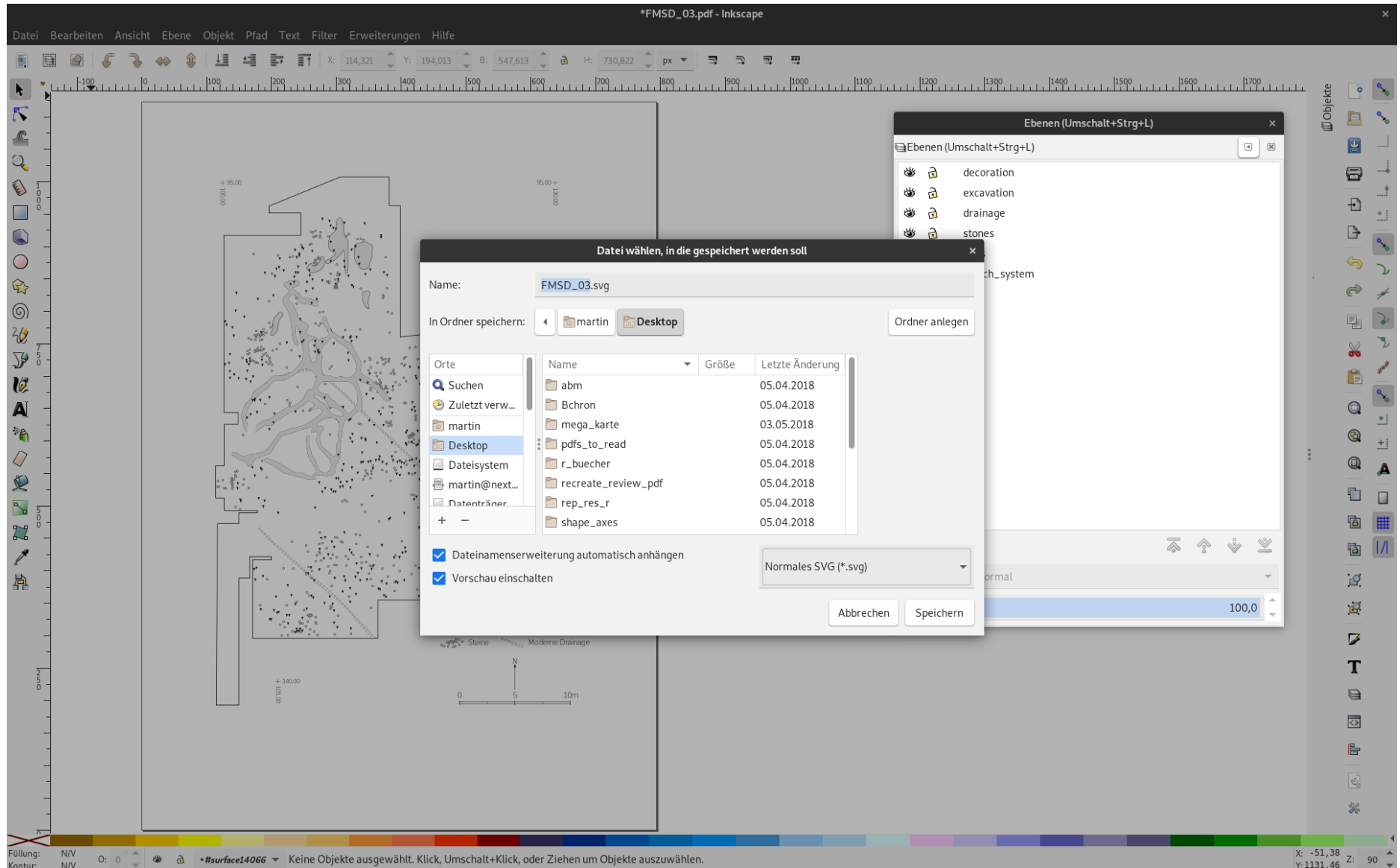
- Variety of proprietary and open-source software for producing vector images, none of which is recommended for long-term - archiving:
  - Coreldraw (.cdr); Adobe Illustrator (.ai); OpenOffice Draw (.odg).
- Think of purpose of vector files and the output.

Illustrations:

- Save output in high quality TIFF or PNG format.

Files with important vector information:

- Document layer conventions
- Export as SVG file (Scalable Vector Graphics)
- PDF files also holds vector data.



# Spreadsheets and Databases: Overview

**Spreadsheets:** Designed on accounting worksheets, primarily for ordering numerical data, performing calculations, and produce charts and figures from data and calculations.

**Databases:** Designed to store a wide variety of data (numerical, text, images) and provide complex search and reporting on these data.

## What is important?

- Data values themselves
- Structure of the tables/sheets used to store
- Structure of relationships between tables in database

# Spreadsheets and Databases

## Data Management

### **Data consistency**

- Standardised data entry is essential.
- Methods for controlled data entry recommended.
- File and field name and codes need to be documented in separate file.
- Document relationships of database tables (screen shot as jpeg)

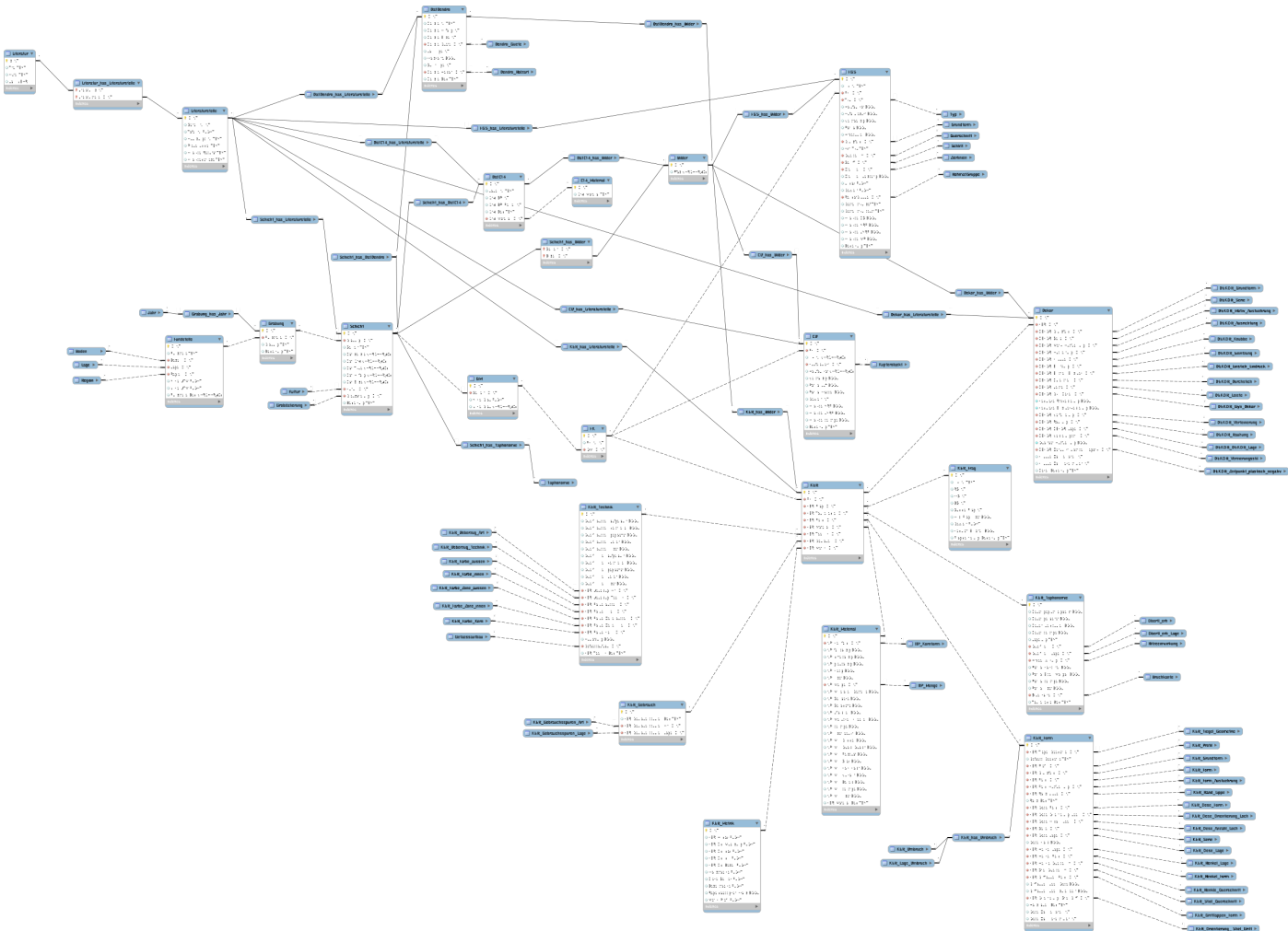
### **Embedded objects**

- Embedded objects (images, charts, figures) stored separately.
- Document analysis/search procedures from which figures are produced.
- Embedded objects removed from final archived file.

### **Non-data content (presentation formatting)**

- Document formatting of tabular data (fonts, colours, cell borders, etc).
- Document data input forms and search query results ('reports').

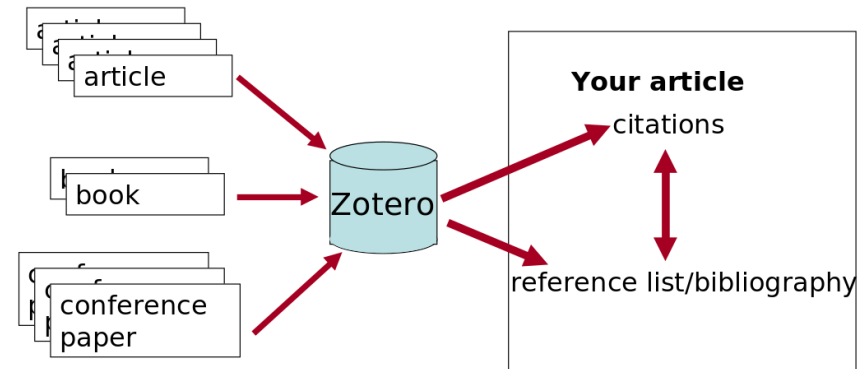
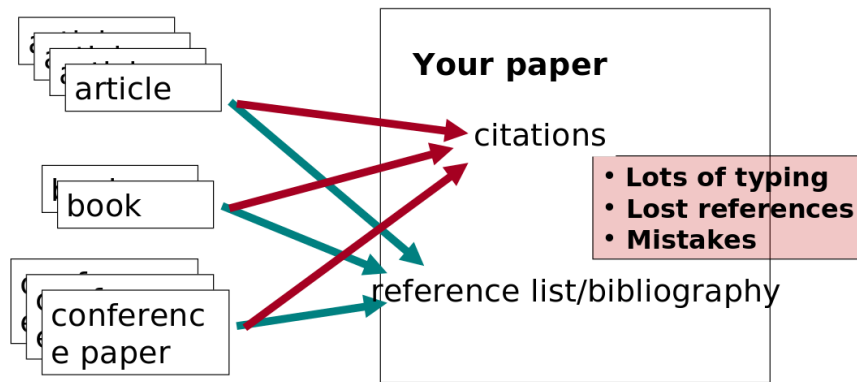
# Database structure (MET-Project)



# Reference management tool

- helps you manage your bibliographic resources and integrates with your word processing software - to generate in-text citations and bibliographies
- saves bibliographic data of information resources from online databases.
- makes it easier to automatically switch to different citation styles (e.g. Harvard to MLA or APA)
- syncs references across multiple devices such as desktops, laptops and tablets – so that you can access the references from any device

Your topic:





# Zotero

Zotero is a free (open source) reference management tool that

- recognizes bibliographic information on books, journal articles, and other resources from websites and databases and extracts metadata from these sources.
- stores related citations, PDFs, files, images, and links in your library
- creates a citations and bibliographies in Word or OpenOffice documents
- publishes shared libraries for your organization, class, or research group

Is available for Mac, Windows & Linux operating systems

To download and install, go to: <http://www.zotero.org>

# Zotero Interface

The Zotero window has three columns that represent different levels of specificity for your references:

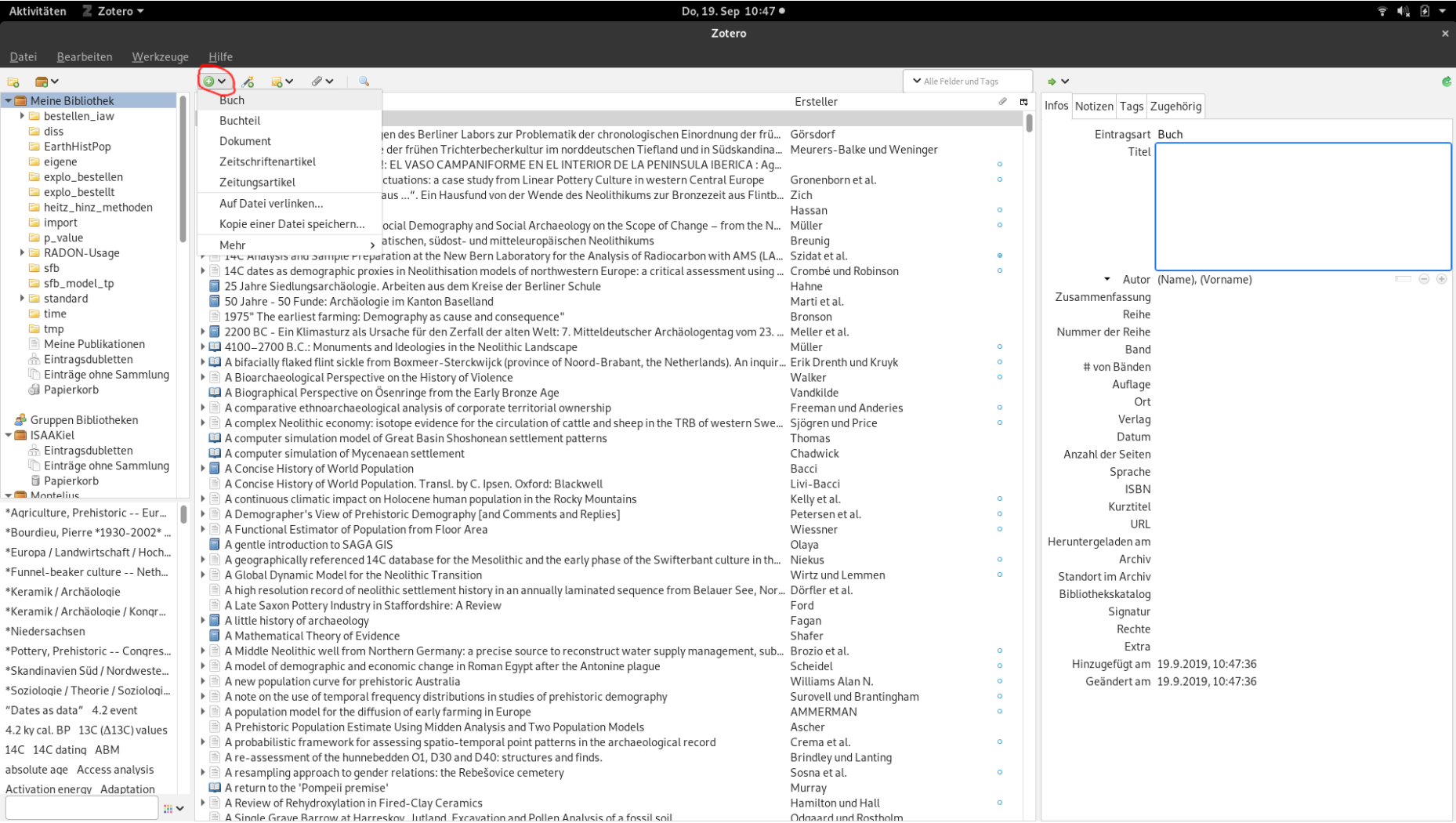
- The left column shows your Zotero libraries and collections as well as 'tags' (keywords)
- The middle column displays items contained within the selected library
- The right column displays information about a selected item.

# Adding citations to Zotero

You can add citations to your library in several ways:

- Manually enter citation information
- Add an item by identifier (ISBN, DOI or PMID)
- Import citation information for
  - a single item
  - multiple items
- automatically from web pages

# Manually enter citation information



# Add an item by identifier (ISBN, DOI or PMID)

The image shows a dual-screen setup. On the left is the Zotero application window, and on the right is a web browser window.

**Zotero Application:**

- The top menu bar includes 'Datei', 'Bearbeiten', 'Werkzeuge', and 'Hilfe'.
- The left sidebar shows a tree view of the 'Meine Bibliothek' (My Library) with various folders like 'bestellen\_iaw', 'diss', 'EarthHistPop', etc.
- The main pane is titled 'Geben Sie ISBNs, DOIs, PMIDs oder arXiv IDs ein, um diese zu Ihrer Bibliothek hinzuzufügen:' (Enter ISBNs, DOIs, PMIDs or arXiv IDs to add them to your library:). A text input field contains '10.12766/jna.2003.4'.
- Below the input field, a list of search results is displayed, showing titles, authors, and other metadata for various publications.

**Web Browser:**

- The address bar shows the URL: [www.jna.uni-kiel.de/index.php/jna/article/view/4](http://www.jna.uni-kiel.de/index.php/jna/article/view/4).
- The page title is 'Das Neolithikum in der Schweiz' (The Neolithic in Switzerland).
- The page features a navigation menu on the left with links: 'Current', 'About', 'Issues', 'Editorial Team', 'Submissions', and 'Search'.
- The main content area displays the article title 'Das Neolithikum in der Schweiz' by Albert Hafner and Peter J. Suter.
- The DOI is highlighted: <https://doi.org/10.12766/jna.2003.4>.
- The publication date is 'Published Feb 23, 2012'.
- Below the title, there are buttons for 'PDF (Deutsch)', 'Tafelteil (Deutsch)', and 'Fundortdaten (Deutsch)'.
- At the bottom, there are buttons for 'Dendro-Daten (Deutsch)' and '14C-Daten (Deutsch)'.
- The abstract text is visible, starting with 'Neolithic research in Switzerland is inseparably bound with its numerous lake side settlements...'.

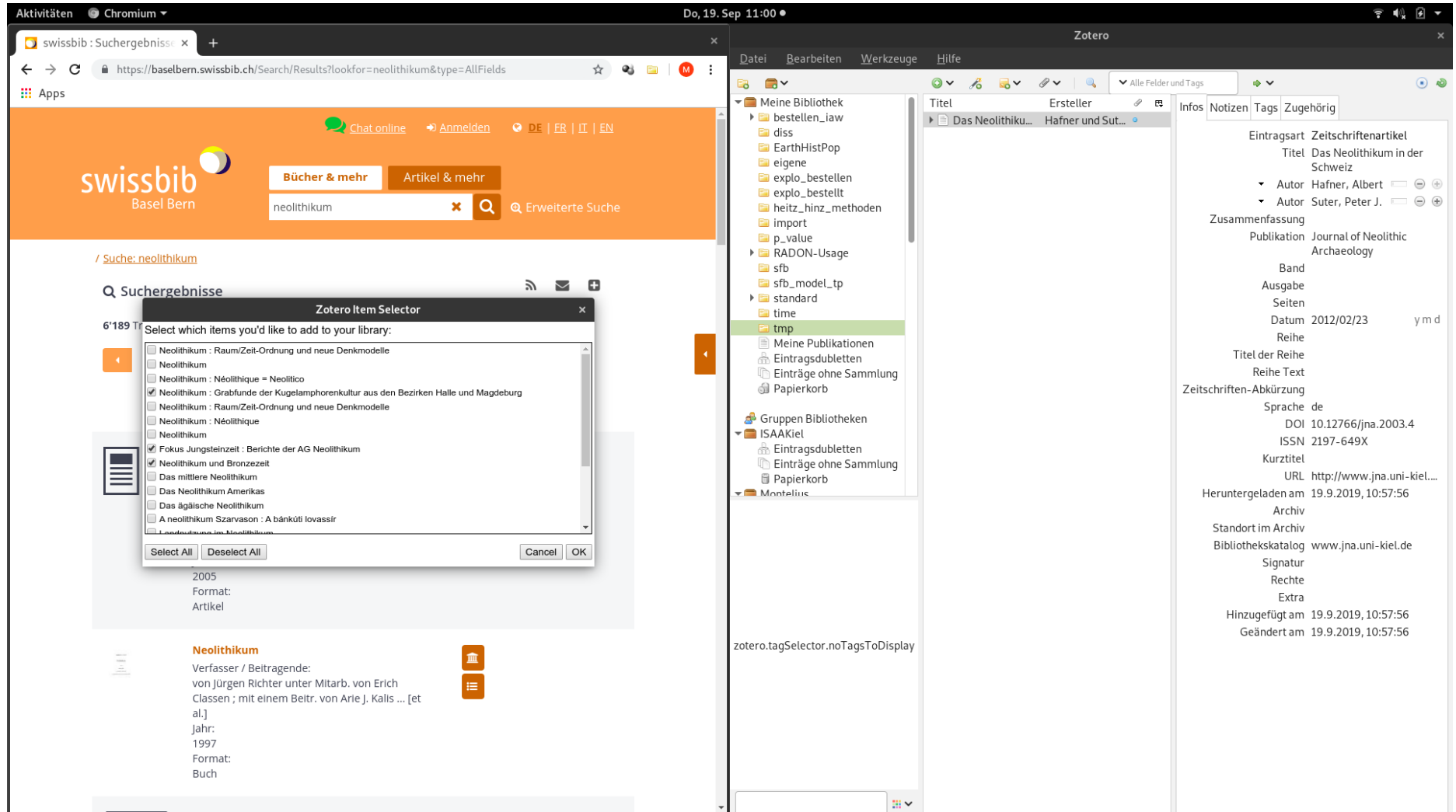
# Import single reference from Web

The screenshot shows a web browser window displaying the article "Das Neolithikum in der Schweiz" by Albert Hafner and Peter J. Suter. A red circle highlights the address bar, and a "Speichern nach" (Save as) dialog box is open, showing the file name "Das Neolithikum in der Schweiz" and the location "tmp".

The Zotero application window is open, showing the "Meine Bibliothek" (My Library) pane on the left, which lists various folders and files. The "Titel" (Title) pane on the right displays the metadata for the selected reference, including the title "Das Neolithikum in der Schweiz", authors "Hafner, Albert" and "Suter, Peter J.", and the publication "Journal of Neolithic Archaeology".




The article text is visible in the browser window, starting with the abstract: "Neolithic research in Switzerland is inseparably bound with its numerous lake side settlements. They were already discovered in the middle of the 19th century and became famous as „Pfahlbauten“. The concept of Neolithic Cultures that was developed in the 1930's by Emil Vogt was undoubtedly influenced by contemporaneous ethnological thinking („Kulturkreislehre“). In his publications Vogt expresses the conviction that „Neolithic Cultures“ are identical with population groups, tribes or peoples. As modern methods of absolute datings did not exist at that time his observations were dependant on typological comparisons. Most importantly, he could not differentiate chronological gaps of several hundred years. Without natural scientific dating methods and relying solely on topography he developed the view that the evolution of ceramics in a region can only be explained by the immigration of foreign people. His proposed chronological chart, which was updated in the 1960/70's, is still in use today. However, in our opinion it is antiquated. During the last two decades dendrochronology and C14 has offered a large number of absolute datings for the Neolithic period. This encourages us to follow new approaches of interpretation. Our concept is based on the well dated material of lake side settlements from Switzerland and Southern Germany. We

# Import multiple references from Web



# Types of Imports

The Save to Zotero icon at the right-end of the URL (address) bar of the browser changes depending on the type of item Zotero identifies on the web page. Here are the most common icons:

Icon	Resemblance	Depicts
	Lined sheet	Journal article
	Blue book	Book
	Manila folder	Multiple items on a search results page (e.g. PubMed or Google Scholar)



# Using Zotero in your word processor

First Citation, Select style

# Using Zotero in your word processor

Further Citations

# Using Zotero in your word processor

## Creating Reference List

# Registering for a Free Zotero Account

Registering for a free Zotero account is useful. It enables you to:

- sync and access your Zotero online from anywhere in the world
- save up to 300MB of content (including PDFs) online
- restore your backed up content to any new Zotero application installation
- participate in Zotero groups and share references
- Zotero group "Montelius"

# Any questions?

You might find the course material (including the presentations) at

<https://berncodalab.github.io/caa>

You can contact me at

[martin.hinz@iaw.unibe.ch](mailto:martin.hinz@iaw.unibe.ch)

Sources for the slides:

'Working with Digital Data' (Lindsay Lloyd-Smith 2011)

'Optimizing your Research Data Management' (Anna Keller, Jennifer Morger & Gero Schreier; Open Science Team Uni Bern)

'Reference Management Software Tools Zotero - Open Source (Module 12)' (research4life)