STADT
GESCHICHTE
BASEL

Digital Longevity

Learnings from the (Digital) History project Stadt.Geschichte.Basel

Dr. sc. Moritz Mähr

Overview

- What is digital longevity?
- Why is digital longevity important?
- How to achieve digital longevity?
  - Ending Principles
  - User Centered Design
- What we do at Stadt.Geschichte.Basel?
  - Research Data Management
  - Public History
- Best practices, Tools, and Workflows

What is digital longevity?

Future archeologists at a dig site, uncovering 20th-century digital artifacts.

What is digital longevity?

Digital longevity, also known as digital preservation, refers to the methods and practices used to ensure that digital information (such as research data) remains findable, accessible, interoperable and reusable (FAIR) over long periods of time, often spanning decades or even centuries.

FAIR Principles in Research Data Management

- **Findable:** Easy to find by both humans and computers, with unique identifiers and machine-readable metadata.
- **Accessible:** Clear guidelines on how to access, including authentication and authorization.
- **Interoperable:** Compatible with other data, applications, and workflows, using standardized formats.
- **Reusable:** Well-described for replication in various settings, with clear usage licenses and provenance.

✅ Enhance data sharing and reusability in scientific research.

Why is digital longevity important?

- Simple answer: the funding agency says so 👉
- The more complex answer:
  - making sure research data is FAIR and stored is expensive and time consuming 💸
  - the earlier in the project you plan and implement it, the less expensive it is 📉

Why is digital longevity important?

“In a print ecosystem many immutable copies of an object are distributed globally and are curated by a network of organizations. The key to maintaining usable analogue artefacts is to ensure that they stay the same over time. Digital objects, in contrast, require active management in order to continue to be easily accessible over time. Software requires patches and updates. File formats can become quickly obsolete as the environments around them change.” (Goddard and Seeman 2020, 38)

Why is digital longevity important?

“It is commonly assumed that storage media will be the most expensive aspect of digital preservation. It’s much more likely, however, that the major cost centre will be the need for human intervention and expert judgement in processes like transforming and migrating objects and metadata over time and in devising legal frameworks for rights management.” (Goddard and Seeman 2020, 39)

How to achieve digital longevity?

- Ending Principles
- User Centered Design

Ending Principles

1. Data Management
2. Documentation
3. Processing Standards
4. Product Development
5. Release Management

Adapted from Endings Principles for Digital Longevity (Version 2.2.1, 2023-03-03).

1. Data Management

Store data in open, processable formats with version control and ongoing validation.

2. Documentation

Clearly document data models and intellectual property, aiming for open licensing where possible.

3. Processing Standards

Write and maintain processing code with relentless validation, continuous integration, and code contingency following data and product principles.

4. Product Development

Create web products without server-side dependencies, using standard, redundant, and independently functional technologies, allowing for graceful failure.

5. Release Management

Plan periodic, coherent, and complete product releases, with clear identification and persistent URLs for citation and accessibility.

Endings Principles

A research initiative by the University of Victoria Library

- **Challenge:** Avoiding a ‘digital dark age’ in Digital Humanities (DH).
- **Context:** 15 years in SSHRC and CFI-funded DH projects. Thousands of digital projects at risk due to limited lifespan (10 years) and evolving digital technologies.
- **Multidisciplinary Team:**
  - **Composition:** Research faculty, programmers, digital librarians.
  - **Expertise:** Geohumanities, multimedia markup, language documentation, textual editing, digital publishing, text encoding.

Endings Principles

- **Research Focus:**
  - Bringing DH projects to closure.
  - Preserving dynamic features of projects.
  - Identifying suitable archival repositories.
- **Goal:** Develop practical strategies for concluding, preserving, and maintaining usability of DH projects.

User Centered Design

1. Understanding User Needs
2. Iterative Design and Testing
3. Multidisciplinary Approach
4. Evaluating Designs Against User Requirements
5. Accessibility and Inclusivity
6. User Feedback

1. Understanding User Needs

Conducting research to understand the users' behaviors, motivations, and objectives.

2. Iterative Design and Testing

Developing prototypes or designs, testing them with real users, and then refining them based on feedback. This cycle repeats until a user-friendly solution is achieved.

3. Multidisciplinary Approach

Involving not just designers, but also usability experts, engineers, developers, and the users themselves.

4. Evaluating Designs Against User Requirements

Ensuring that the final product meets the actual needs of its intended users.

5. Accessibility and Inclusivity

Designing products that are accessible to all users, including those with disabilities.

6. User Feedback

Regularly obtaining and incorporating feedback from users to improve the design.

How do the Ending Principles and User Centered Design play together?

- Ending Principles cover technical aspects of digital longevity.
- User Centered Design covers the human aspects (such as acceptance and adoption) of digital longevity.

Stadt.Geschichte.Basel

- Major historical research project at the University of Basel from 2017 to 2025.
- More than 80 researchers studying the history of Basel from the beginnings of settlement to the present day.
- Funded with over CHF 9 million from the Canton of Basel-Stadt, the Lottery Fund and private donors.
- The project has a team for research, data management and public history (1 project manager, 3 data stewards, 2 public outreach).
- The outputs of the project are 10 printed volumes and a digital portal.

Mission and goals

- An online portal. It is designed as a virtual repository that provides public access to a wide range of data on Basel's history (...). The portal will be designed and organizationally integrated in such a way that it can be continued and continuously updated after the end of the eight-year project period.
- The continuous, public visibility and the dialogue with future users of a city history are important concerns of the project.

Excerpt from the contract between the Canton of Basel-Stadt and the Foundation Stadt.Geschichte.Basel dated 31.7.17, translation MM

Work packages

1. Securing the research data of the urban history project; **(Research Data Management)**
2. Development of showcases for the preparation and communication of relevant data and media on urban history; **(Public History)**
3. Clarification of metasearch functions and integration and networking of own and external content.

Excerpt from application “Stadt.Geschichte.Basel digital” of 27.1.2021, translation MM

Research Data Management

- Research Data Management (RDM) refers to the organisation, storage and preservation of data created during a research project. It covers initial planning, day-to-day processes and long-term archiving and sharing.
- Data here is used as a generic term to encompass all digital objects. RDM is a vital part of enabling reproducible research. RDM ensures efficiency in research workflows, and also greater reach and impact, as data become FAIR (Findable, Accessible, Interoperable and Reusable).

Adapted from The Turing Way: A handbook for reproducible, ethical and collaborative research. (27-07-2022)

Research data management in the historical sciences

What is missing [in the historical sciences], however, is an awareness that we are **not only working with digital or digitized texts, images, object descriptions, etc., but also with data.** Also, the subject-specific recommendations for dealing with research data are available for subjects such as biology and psychology, or even for educational science, but not for historical science. (Hiltmann 2018 Translation MM)

Historical research data?

While the narrower term actually included only data generated in the concrete research process of the individual projects, such as the text annotations or metadata, the broader term also included the sources underlying the historical research process as a whole - as they are scientifically processed by libraries, archives, and museums and thus made available for potential use by researchers. (Hiltmann 2018 Translation MM)

Research data at Stadt.Geschichte.Basel (Mähr 2022)

- The sources are text data (93%), image data (90%) and tables (60%). Video, audio, and geospatial data are used by less than 15%.
- Data is used to substantiate statements (100%), create maps (40%), analyze networks (27%), and create time series and quantitative analysis (23%).
- PDF software (88%), spreadsheet programs (58%), and GIS programs (17%) are used to annotate sources.
- Sources are recorded with word processing programs (93%), on paper (60%), and with literature management software (40%).

Research Data Management at Stadt.Geschichte.Basel

- Collecting research data is costly and places high demands on all parties involved. Therefore, it should be simple, practical and accessible.
- Three data stewards have been hired for the four most important data types: Images and figures, maps and georeferences, tables and statistics, and bibliographic references.
- Project-wide solutions have been defined for common data types. For little-used data types, specific solutions are provided upon request.
- The data are collected together with the manuscripts for the printed volume. This minimizes the workload for authors and centralizes data quality control.
- Free training on digital working tools and research practices is offered, based on the needs of the researchers.

From research data to public history?

- Research data must be prepared (or produced) for a very specific audience.
- Research data are part of a data ecosystem. Redundancy in collection must be avoided. Existing data pools should be tapped or replenished.
- The public cannot be informed directly with research data. It needs a circuitous route through intermediaries.
- Therefore we cooperate with different intermediaries (memory institutions etc.).

Insights from surveys with the ecosystem

- There is a need for target group-oriented digital and analog services and products for the mediation of research data
- Collaboration with existing providers, integration of existing services and data
- Identifying and addressing gaps (data and services) within the ecosystem
- Small projects, rapid prototyping, and ongoing monitoring of success (e.g., with surveys)
- Low technical and legal barriers to entry increase the chance of data reuse

Best practices (Mähr 2023)

1. Durability by Design
2. Openness and Transparency
3. Respect for User Community
4. Recognition of Complexity
5. Sustainable Resourcing

1. Durability by Design

- Focus on long-term viability of digital resources.
- Usage of open, non-proprietary formats for data and metadata.
- Continuous updates and maintenance of the online ressources post-project.

2. Openness and Transparency

- Commitment to making project work accessible and transparent.
- Engagement with diverse stakeholders to meet community needs.

3. Respect for User Community

- Regular surveys to understand user needs.
- Inclusion of users in development and as data producers.
- Data Stewards' role in managing researchers' needs and offering support.

4. Recognition of Complexity

- Acknowledging the challenges of digital preservation.
- Dedicated Data Stewards managing digital resources.
- Ongoing training to keep pace with research data management advances.

5. Sustainable Resourcing

- Allocating resources specifically for digital preservation.
- Partnerships with institutions for long-term digital resource preservation.

What we are currently working on

- Research Data Platform
- Portal
- Data Stories
- Scientific illustrations (georeferenced maps, statistical analyses, etc.)
- Guidelines for inclusive annotation of historical data
- Open Data Workflows

Research Data Platform

- forschung.stadtgeschichtebasel.ch (clickable prototype)
- Built with CollectionBuilder-CSV, a standalone template for digital collections, exhibit websites, and teaching websites
- Open Source
- Minimal and static
- Simple and stupid (Jekyll, CSV, Markdown)
- Actively maintained by the University of Idaho Library (and us)
- github.com/CollectionBuilder/collectionbuilder-csv

Portal

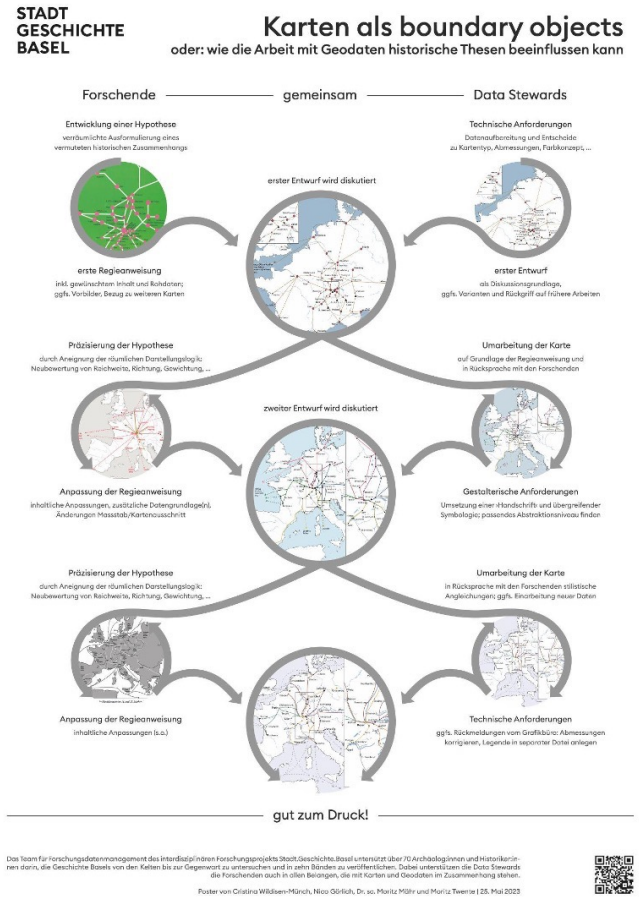
- stadtgeschichtebasel.ch
- Built with SvelteKit and SkeletonUI
- Using a https://sgb.hypotheses.org/ as a headless CMS for the blog
- Using the XML API by https://www.agendabasel.ch/ for the agenda
- Open source github.com/Stadt-Geschichte-Basel/stadtgeschichtebasel.ch
- Minimal and static

Data Stories

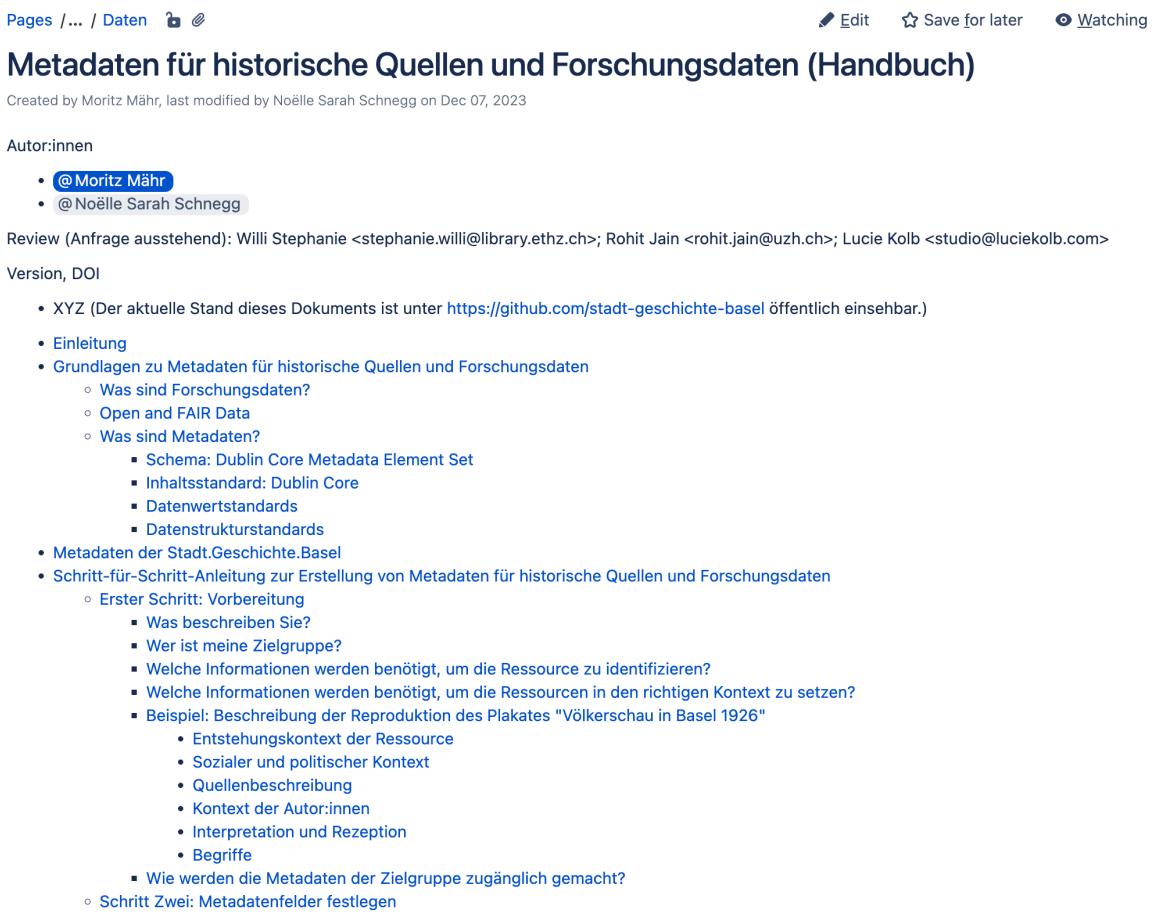
[Das Kontrollbüro und die Wundermaschine]

Weitere Stories befinden sich in der Entwicklung.

Scientific illustrations (Wildisen-Münch et al. 2023)



Guidelines for inclusive annotation of historical data



Pages /... / Daten Edit Save for later Watching

Metadaten für historische Quellen und Forschungsdaten (Handbuch)

Created by Moritz Mähr, last modified by Noëlle Sarah Schnegg on Dec 07, 2023

Autor:innen

- @Moritz Mähr
- @Noëlle Sarah Schnegg

Review (Anfrage ausstehend): Willi Stephanie <stephanie.willi@library.ethz.ch>; Rohit Jain <rohit.jain@uzh.ch>; Lucie Kolb <studio@luciekolb.com>

Version, DOI

- XYZ (Der aktuelle Stand dieses Dokuments ist unter https://github.com/stadt-geschichte-basel öffentlich einsehbar.)

Open Research Data Workflows

open-research-data-template

GitHub template for FAIR and open research data

View the Project on GitHub
maehr/open-research-data-template

Download ZIP File | Download TAR Ball | View On GitHub

This project is maintained by maehr

Hosted on GitHub Pages — Theme by orderedlist

GitHub template for FAIR and open research data

This GitHub template is independent of the (research) data and its format. It follows the best practices for open research data as outlined in The Turing Way. It uses GitHub Actions to manage releases, issues, and pull requests, GitHub Pages for documentation, and Zenodo for long-term archiving.

issues 5 open | forks 0 | stars 16 | license CC-BY-4.0

Why use a template (even for small datasets)?

- Share your open research data with others
- Write better documentation for yourself and the community
- Write more consistent code and encourage collaboration
- Increase security
- Follow accepted ethics

How this template helps you

Open research data

- Citeable via CITATION.cff and DOI
- Automatic long-term archiving via Zenodo
- Licensed under a non-restrictive CC BY 4.0 license according to The Turing Way
- Templates for reporting data issues using a custom template

Documentation

- README.md according to www.makeareadme.com and The Turing Way
- CHANGELOG.md according to keepachangelog.com
- Automated CHANGELOG.md via git-cliff
- package.json via npm docs
- Accessible documentation via gh-pages

maehr.github.io/open-research-data-template/

Thank you for your attention!

Bibliography

Goddard, Lisa, and Dean Seeman. 2020. “Negotiating Sustainability: Building Digital Humanities Projects That Last.” In *Doing More Digital Humanities: Open Approaches to Creation, Growth, and Development*, edited by Constance Crompton, Richard J. Lane, and Raymond George Siemens. London ; New York, NY: Routledge/Taylor & Francis Group.

Hiltmann, Torsten. 2018. “Forschungsdaten in der (digitalen) Geschichtswissenschaft. Warum sie wichtig sind und wir gemeinsame Standards brauchen.” Billet. *Digitale Geschichtswissenschaft* (blog). September 17, 2018. https://digigw.hypotheses.org/2622.

Mähr, Moritz. 2022. “Research Data Management in (Public) History.” Key note, Istituto Svizzero di Roma, June 15. https://doi.org/10.5281/zenodo.6637118.

———. 2023. “Embracing endings: Principles for digital longevity and their importance for research software engineers.” Blog. DHTech. August 7, 2023. https://dh-tech.github.io/blog/2023-08-07-ending-principles/.

Wildisen-Münch, Cristina, Nico Görlich, Moritz Mähr, and Moritz Twente. 2023. “Karten als ‘boundary objects’ oder wie man mit Geodaten historische Thesen bildet.” Berlin, May 23. https://doi.org/10.5281/zenodo.7960745.