Archaeology Data Infrastructures

Data reuse potentials and limitations to modelling settlement systems (…)

Petr Pajdla

2022-11-12

Table of contents

# Preface

|  |  |
| --- | --- |
|  | **Warning**  This is a website for the **work-in-progress** PhD thesis of mine. It is **not** intended to be read by anyone except me *(and maybe few other people)* yet. If you do flick through it anyway, consider yourself warned. It might be messy at some places and will definitely undergo serious rewriting. |

|  |  |
| --- | --- |
|  | **Note**  This work can be read online at <https://petrpajdla.github.io/dataInfrastructures/>. The source repository is on GitHub at <https://github.com/petrpajdla/dataInfrastructures/>. |

This document is created in an open-source [Quarto](https://quarto.org/) scientific and technical publishing system. You might be asking why is it published and written like this even if it is not intended for any audiences except myself yet. I have no answer to this. One evening I simply decided to give *Quarto* publishing a try and set this whole thing up in less then an hour or so.

### Notes on writing

This note is written mostly for a future me, in case I need to set up the working environment again on a different machine and to serve as a memo if I forget how to continue.

As of November 2022, this is written on [Archlabs](https://archlabslinux.com/) *GNU/Linux* machine, mostly in [Visual Studio Code](https://code.visualstudio.com/) editor and sometimes in [RStudio](https://posit.co/). Changes are trackeg with *Git* and a remote repository is on *GitHub* (see the note above), same as the rendered website. The rendered version of the manuscript is in the branch gh-pages. See a guide on how to set this up [here](https://quarto.org/docs/publishing/github-pages.html). The online version is published with this command:

**Terminal**

quarto publish gh-pages

In my point of view, there are numerous advantages to scientific writing in this manner over traditional *Office*-based approach. A non-exhaustive list of why to do scientific writing this way is below.

* **Plain text**  
  Writing in plain text enhanced with a simple *Markdown* syntax and some *Quarto* elements is great because from one source document, a *.pdf*, *.html*, *.docx* (and probably more) document formats can be rendered using [pandoc](https://pandoc.org/).
* **Version control**  
  Tracking changes using *git* is easily implemented when writing in a plain text. Keeping track of any changes in the manuscript is obviously crucial for any later revisions etc.
* **Simple citation management**  
  Bibliography is organized using [Zotero](https://www.zotero.org/) with [Better BibTeX](https://retorque.re/zotero-better-bibtex/) extension which is used to export (and keep updated) necessary collections in a parent folder of the manuscript as *.bib* files. My *Zotero* library is [here](https://www.zotero.org/knytt). To format the citations, a citation style of the *Journal of Computer Applications in Archaeology* is used (.csl file was obtained [here](https://www.zotero.org/styles/)).
* **Embedded code**  
  Code blocks (and the associated results) can be easily embedded in the text. My language of choice is *R*. For more information on reproducibility see Marwick (2017) and Marwick, Boettiger and Mullen (2018).

# Introduction

This is a book created from markdown and executable code.

See Zhang, Zhao and Ventrella (2018) for additional discussion of literate programming.

# 1. Theory and Method

## 1.1 Archaeology as theory- and/or data-driven science

(based on TAG Brno 2021 talk)

## 1.2 Theorizing data

Defining archaeological data, micro- to macro-scales;

## 1.3 Methodological Approaches

Review of current approaches: Spatial and/or Landscape archaeology, Macroarchaeology,Big data archaeology etc.

# 2. Data

Sources of (archaeology) data in the Czech Republic, an overview:

Data models, datafication of past reality, simple vs complex data models; Assessing findability, accessibility, interoperability, and reusability (FAIR) principles; Cultural heritage management data vs research data domains; Archaeological information system of the Czech Republic (AIS CR) as the main data infrastructure.

# 3. Discussion

# 4. Summary

In summary, this book has no content whatsoever.

# References

Marwick, B. 2017 Computational Reproducibility in Archaeological Research: Basic Principles and a Case Study of Their Implementation. *Journal of Archaeological Method and Theory* 24(2): 424–450. DOI: https://doi.org/[10.1007/s10816-015-9272-9](https://doi.org/10.1007/s10816-015-9272-9).

Marwick, B, Boettiger, C and Mullen, L. 2018 Packaging Data Analytical Work Reproducibly Using R (and Friends). *The American Statistician* 72(1): 80–88. DOI: https://doi.org/[10.1080/00031305.2017.1375986](https://doi.org/10.1080/00031305.2017.1375986).

Zhang, S, Zhao, B and Ventrella, J. 2018 Towards an Archaeological-Ethnographic Approach to Big Data: Rethinking Data Veracity. *Ethnographic Praxis in Industry Conference Proceedings* 2018(1): 62–85. DOI: https://doi.org/[10.1111/1559-8918.2018.01197](https://doi.org/10.1111/1559-8918.2018.01197).