EVOLVING TOWARDS AN ERA OF OPEN RESEARCH

Scriberia

Working towards an open, collaborative and reproducible data culture in archaeology

Esther Plomp

she/her / e.plomp@tudelft.nl / @toothFAIRy

0000-0003-3625-1357

Data Steward @TU Delft, Faculty of Applied Sciences, NL

Open Research Ambassador @IsoArcH

Project Member @The Turing Way

Mentor @Open Life Science

Team Member @Open Research Calendar

Open Data Expert @TOPS/NASA & OpenSciency contributor

Emma Karoune · Alice Toso · Kirstie Whitaker · Corina Logan · Paula Martinez Lavanchy · Lena Karvoskaya · Yasemin Türkyilmaz-van der Velden · Arielle Bennett · Yo Yehudi · Cassandra Gould van Praag · Rowland Mosbergen · Marta Teperek · Malvika Sharan · Joshua Peterson · Emmy Tsang · Laura Acion · Bjørn Peare Bartholdy · Kevin Salesse · Yanina Bellini Saibene · Emma Henderson · Chris Stantis · Corinne Hofman · Lisette Kootker · Paulien Herder · Chris Cheung · Ben Marwick · My manager: Marja van den Bergh · Frederique Belliard · Angelica Maineri · Hao Ye · Sabrina Meindlhumer · Anne Lee Steele · Saranjeet Kaur Bhogal · Heidi Seibold

It takes a (n open science) community

Content

My story

What led me to working more openly?

Why?

What are the benefits of working more openly and reproducibly? For you and for the wider discipline.

How?

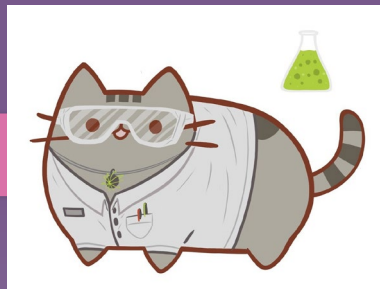Tools and communities that can support you in making your research more openly available.

Joining Open Science communities (The Turing Way)

Finishing PhD

Start Archaeology studies at Leiden University, the Netherlands

Start PhD research at Vrije Universiteit Amsterdam, the Netherlands

2008

2013

2018

2020

Start position as a Data Steward at TU Delft, the Netherlands

Illustrations by Kelly Stanford

Timeline

Method paper 2017

https://doi.org/10.1039/C7JA00312A

👍 Elaborate description of research protocol

TIMS analysis of neodymium isotopes in human tooth enamel using 10¹³ Ω amplifiers   Check for updates

E. Plomp,[a] I. C. C. von Holstein,[a] J. M. Koornneef,[a] R. J. Smeets,[a] L. Font,[b] J. A. Baart,[cd] T. Forouzanfar[cd] and G. R. Davies[a]

Postprint on EarthArXiv10.31223/osf.io/ky9e3

4 Methodology

4.1 Sample preparation

All teeth were sampled, chemically processed and analysed at the Faculty of Science, Vrije Universiteit Amsterdam. The teeth were leached overnight in 30% H_2O_2 (Sigma-Aldrich Company Ltd) followed by a rinsing step in ultrapure water (Milli-Q) and air dried on a hotplate at 50 °C prior to sampling. The enamel was sampled using a dental micro-drill fitted with a diamond-tipped rotary burr and blade (Minilor Perceuse). The burr and blade were cleaned between sampling with 3 N HNO_3 in an ultrasonic bath for 3 minutes to remove any residual particles and then rinsed with Milli-Q and dried. The innermost layer of enamel (at the enamel–dentine junction) was mechanically removed to ensure optimal enamel–dentine separation. Sample weight ranged from 273–1233 mg, average 515 mg. If a M3 pair was available from a single donor, the enamel from both teeth was combined to increase available sample size.

4.2 Chemical separation

Sample dissolution and Nd chromatographic separation was performed in a class 100 clean laboratory. All PFA laboratory equipment was sub-boiled in pro-analysis quality 7 N HNO_3 and 6 N HCl for two hours each, followed by two leaching steps at 125 °C with (1) double distilled 6.5 N HCl (>5 days) and (2) 7 N HNO_3/12 N HF (>2 days).

Thorough removal of organic matter was essential due to large sample size required to obtain enough Nd for analysis. The enamel was dissolved in 3–6 mL 6.5 N HCl, dried, nitrated, and re-dissolved in 3–6 mL 6.5 N HCl and 0.75–1.5 mL 14.0 N HNO_3, depending on the sample weight. The solution was dried and nitrated before it was dissolved in 10 mL 2.0 N HNO_3 for the extraction of the LREE with TRU-resin using an adapted protocol described in Section 4.2.2. After LREE extraction Nd was separated from the other LREE using Ln-resin (Eichrom Technologies) following standard procedure.[50]

4.2.1 Neodymium isotope concentration.
In order to determine the range of Nd concentrations in human teeth, isotope dilution using a ^{150}Nd enriched spike ($^{150}Nd/^{144}Nd = 142.93$) was performed on a 10% aliquot of a subset of the samples ($n = 23$). Sample aliquots were chromatographically separated following standard TRU-resin column extraction (Eichrom, 150 µL resin, 100–150 µm) and standard Ln-resin columns (Eichrom, 740 µL, 50–100 µm).[50]

4.2.2 Neodymium isotope composition.
The calcium-rich enamel samples of >250 mg required the development of a new Nd separation protocol, to prevent overloading of conventional chromatographic columns and to achieve removal of Ca in the purified Nd fraction, while minimizing the procedural blank because of low Nd concentrations in the samples (<0.1 ppm). Excessive Ca in the tooth enamel samples initially caused poor Nd column yields and low ionisation efficiency in the mass spectrometer. Removal of Ca was accomplished with a novel TRU-resin protocol using a modified column made from Pasteur pipettes (Fig. 2). The TRU-resin volume used depended on the sample size: 0.75 mL resin for samples up to 550 mg and 1.3 mL resin for samples >550 mg.

Column extraction followed the procedure described in Fig. 2. Samples were ultrasonicated for 30 minutes and centrifuged for 4 minutes at 4000 rpm before loaded onto the column. After the first extraction in 8 mL (samples <550 mg) or 14 mL (samples >550 mg) Milli-Q, the REE fraction was collected and

1. REE separation
0.75 or 1.3 mL TRU-resin

Cleaning
6 mL 2 N HF
6 mL Milli-Q
6 mL 2 N HNO_3
6 mL Milli-Q
Precondition
6 mL 2 N HNO_3
Prefraction (25 CV)
19 or 33 mL 2 N HNO_3
   Sample load: 10 mL 2 N HNO_3
   Wash: 9 or 23 mL 2 N HNO_3
REE extraction (10 CV)
8 or 14 mL Milli-Q
Precondition
6 mL 2 N HNO_3
Prefraction
   Sample load: 9.6 or 16.8 mL 2 N HNO_3
   Wash: 9.4 or 16.2 mL 2 N HNO_3
Second REE extraction
8 or 14 mL Milli-Q

2. Nd separation
0.74 mL Ln-resin

Cleaning
4 mL 6 N HNO_3
4 mL 2 N HF
4 mL Milli-Q
4 mL 6-7 N HCl
1 mL 0.165 N HCl
Storage overnight
Cleaning
2 mL 6-7 N HCl
2 mL Milli-Q
Precondition
2 mL 0.165 N HCl
Prefraction
9-11 mL 0.165 N HCl
   Sample load:
   1-2 mL 0.165 N HCl
   Wash:
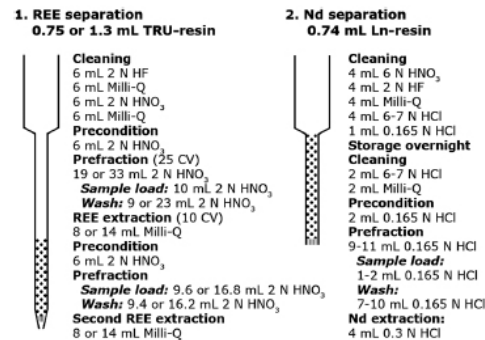   7-10 mL 0.165 N HCl
Nd extraction:
4 mL 0.3 N HCl

Fig. 2   Schematic overview of the chemical separation method for Nd separation for Ca-rich samples. REE from samples and standards were first extracted using TRU-resin employing Pasteur pipettes (35 µm polyethylene frit). Subsequently the Nd was extracted using Ln-resin. CV stands for column volume, which equals the volume of resin used.

Evaluation paper 2019

👎 One sentence descriptions

Evaluation of neodymium isotope analysis of human dental enamel as a provenance indicator using 10^{13} Ω amplifiers (TIMS)

E. Plomp [a] 👤 ✉, I.C.C. von Holstein [a], J.M. Koornneef [a], R.J. Smeets [a], J.A. Baart [b c 1], T. Forouzanfar [b c], G.R. Davies [a]

5. Methods

5.1. Sample preparation

The enamel was sampled, chemically processed and analysed at the Faculty of Science, Vrije Universiteit Amsterdam. Sample preparation and procedures are described in detail in Plomp et al. [19]. The enamel was sampled using a dental micro-drill fitted with a cleaned diamond-tipped rotary burr and blade (Minilor Perceuse). Sample weight for Nd composition in this study ranged from 222 to 1464 mg (average = 733 mg, $n = 20$) of which 1–2% aliquots were taken for Sr analysis for the individuals from South Holland, Limburg, Friesland, the Caribbean, Columbia and Iceland ($n = 25$). If two third molars were available from a single donor, the enamel from both teeth was combined to increase available sample size.

5.2. Chemical separation

Sample dissolution and chromatographic separation was performed in a class 100 clean laboratory. All PFA laboratory equipment was cleaned according to standard procedure [19]. In order to assess the variability introduced by the laboratory procedures a synthetic tooth standard (TSTD) was used [19]. TSTD aliquots were processed on 0.75 and 1.3 mL TRU-resin columns (10 mL, 4 ng Nd, 1000 mg $CaHPO_4$) and Sr columns (0.05 mL, 500 ng Sr, 5 mg $CaHPO_4$).

Protocols.io

- Online platform for protocol management

- Publishing + DOI

https://www.youtube.com/watch?v=84B8P6BAOgM
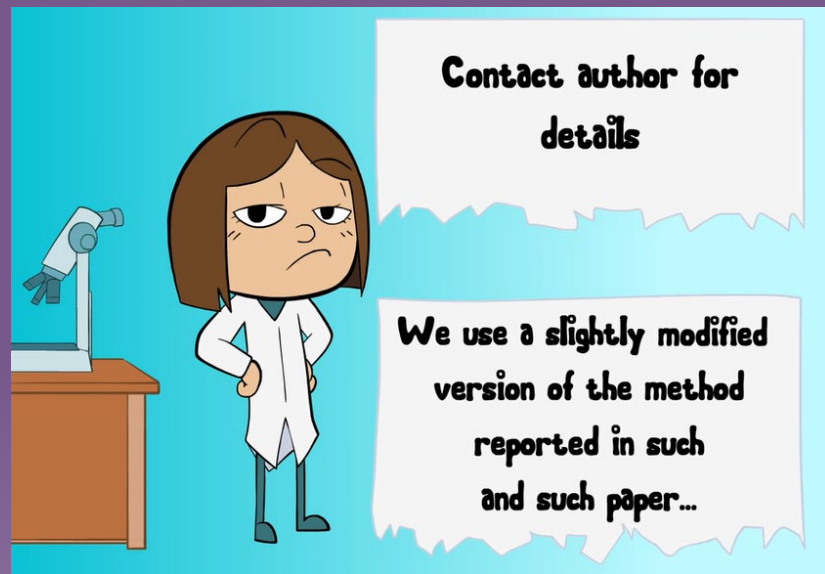
Protocols.io

Sabin, S. & A Fellows Yates, J. (2020) 'Dental Calculus Field-Sampling Protocol (Sabin version) v2' [Protocol] protocols.io. doi:10.17504/protocols.io.bqecmtaw

🌐 Dental Calculus Field-Sampling Protocol (Sabin version) V.2

In 5 collections

DOI
dx.doi.org/10.17504/protocols.io.bqecmtaw

Susanna Sabin[1], James A Fellows Yates[2]

[1]Max Planck Institute for the Science of Human History;
[2]Max Planck Institute for Evolutionary Anthropology

WarinnerGroup   MPI EVA Archaeogenetics

Zandra Fagernäs

VERSION 2
DEC 11, 2020

SHARE

WORKS FOR ME   1

COMMENTS 0

★ BOOKMARK

⇌ RUN

COPY / FORK

MORE ↓

STEPS   GUIDELINES   WARNINGS   FORKS   METADATA   MATERIALS   METRICS

ABSTRACT

This protocol describes how to sample dental calculus from skeletal remains for biomolecular analysis. This protocol is particularly recommended for sampling calculus from teeth attached to a jaw bone, skull, or skeleton. The primary use-case is for DNA and proteomic analysis.

BEFORE STARTING

Decide on a sampling approach e.g. will you pool calculus from multiple teeth or surfaces (increasing yield, but loss of bio-geographic information)

Make sure you have a bin avaliable. You will produce a lot of waste to ensure minimal contamination.

"

Neodymium extraction followed the procedure described in Plomp et al. [19] (available on protocols.io: dx.doi.org/10.17504/protocols.io.xzmfp46)

E. Plomp, R. Smeets, G. Davies, [protocol] Chromatographic separation of strontium isotopes in human dental enamel for thermal ionization mass spectrometry (TIMS) analysis v1, protocols.io, 2020. doi:10.17504/protocols.io.37dgri6

"

"What is your protocol for processing this standard? "

See my published protocol!

Started PhD in 2013, published peer reviewed article in 2017
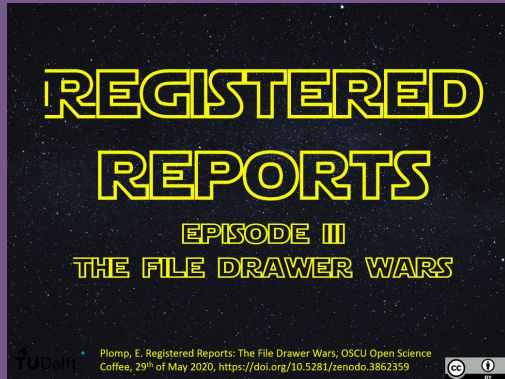
Dutch PhD candidates are paid for 4 years

PhD in 4 - 5, 6, 7

Negative results

- Difficult to publish negative results
- However: negative results = scientific progress

If we had a different research culture, my PhD would have been less stressful...

Negative results

Registered Reports

- Since 2013 (Center for Open Science)

- Publishing format where research question and the quality of methodology are central by conducting peer review prior to data collection

REGISTERED REPORTS

EPISODE III
THE FILE DRAWER WARS

Plomp, E. Registered Reports: The File Drawer Wars, OSCU Open Science Coffee, 29th of May 2020, https://doi.org/10.5281/zenodo.3862359

DEVELOP IDEA → DESIGN STUDY → COLLECT & ANALYZE DATA → WRITE REPORT → PUBLISH REPORT

Stage 1 Peer Review

Stage 2 Peer Review

Results are a 'dead currency' in quality evaluation

Registered Reports: Increased impact of peer review

- Feedback from reviewers at a time that you can still incorporate it

- For example: What if you plan to sample unburned, buried bone for strontium isotope analyses?

"

Furthermore, this work is a demonstration of **why [journal] should consider to accept registered report articles** (https://www.cos.io/initiatives/registered-reports). **Through registered reports the proposal of the work (introduction/methods) is first reviewed and accepted before any analyses are carried out. This could prevent wasting time/samples** from the start, rather than having to point this out when analyses are already carried out and **nothing can be [done].**

Definitions

Open, Reproducible, FAIR, CARE

Open Science

□ ..defined as an **inclusive construct** that combines various movements and practices aiming to make **multilingual scientific knowledge openly available, accessible and reusable** for everyone, to **increase scientific collaborations and sharing of information** for the benefits of science and society, and to open the processes of scientific knowledge creation, evaluation and communication to **societal actors beyond the traditional scientific community**. — UNESCO recommendation on Open Science

Open Science Buffet *

The term Open Science Buffet has been coined by Christina Bergmann in 2019

Open Science
is like a buffet*: take what you can and what benefits you now - come back for more!

- Open Data
- Open Software
- Open Hardware
- Open Publishing
- Open Engagement
- Open Education
- Open Methods
- Open Participation
- Open Evaluation

Reproducibility

▫ When data from the original study can be reanalysed taking the same steps and producing the same or similar results

		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable

https://the-turing-way.netlify.app/reproducible-research/overview/overview-definitions.html

FAIR principles

FINDABLE

Archive your data in a **data repository** with metadata and a **persistent identifier**

ACCESSIBLE

Determine what should be shared

Ensure there is an access procedure in place

≠ open!

Use **open/common formats** and languages

Consistent vocabulary

Metadata standards

INTEROPERABLE

Apply a **licence** to specify how others can reuse your data

Documentation

REUSABLE

FAIR? (Presentation slides) / GO FAIR overview /
Wilkinson et al. 2016 (Paper on FAIR)/ The Turing Way: FAIR (general overview)

21

CARE principles

The **CARE principles** facilitate Indigenous control in data governance and reuse, promoting equitable participation (Carroll et al. 2020)

Relevant for many populations (privacy, future use, reuse, stewardship concerns)

Can be used as a standard in crafting policies on data acquired about communities or populations

Consultation with communities is expected (Turner & Mulligan 2019)
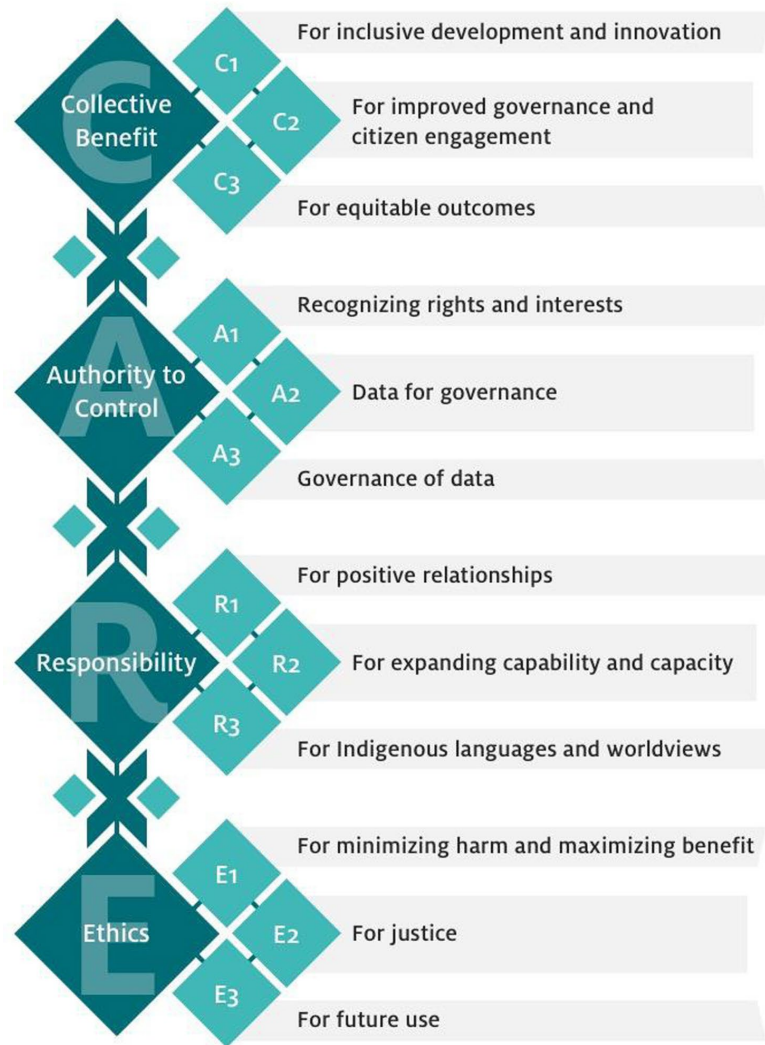
Open: Multilingual scientific knowledge, openly available, involving stakeholders
Reproducible: Transparent workflows
FAIR: Technical aspects: Data repository, documentation, license
CARE: Human aspects: Ethical research

23

Why?

Open, Collaborative, Reproducible

Benefits

- **Preservation** of research outputs
    - Can you still use data/software when people leave?
- **Recognition** & control over all research outputs (data, code, protocols)
- **Collaboration**
- Increases **quality** of and trust in scientific practice (increased transparency)
- Equitable access to resources
- Cost/time efficient (after a learning curve)

Archaeology

1. Archaeology is destructive
2. CARE/Equal access to knowledge generated by these remains

Removing Barriers to Reproducible Research in Archaeology https://doi.org/10.5281/zenodo.7320029

Colleen Morgan
@clmorgan@archaeo.social

Archaeology is great because you not only get to curse past people for doing weird things you don't understand, you also get to curse your past self for excavating and recording these weird things badly in a way that you don't currently understand.

Damn you, past Colleen.

03 mei 2023 07:03 · ⊕ · Web · ⟲ 2 · ★ 6

Required

Funders:
- ▫ European Commission
- ▫ Wellcome, NWO (Dutch Research Council)

Institutions:
- ▫ TU Delft: all individuals in research should follow FAIR principles

Journals:

"The American Journal of Biological Anthropology expects you to archive all the data from which your published results are derived in a public repository."
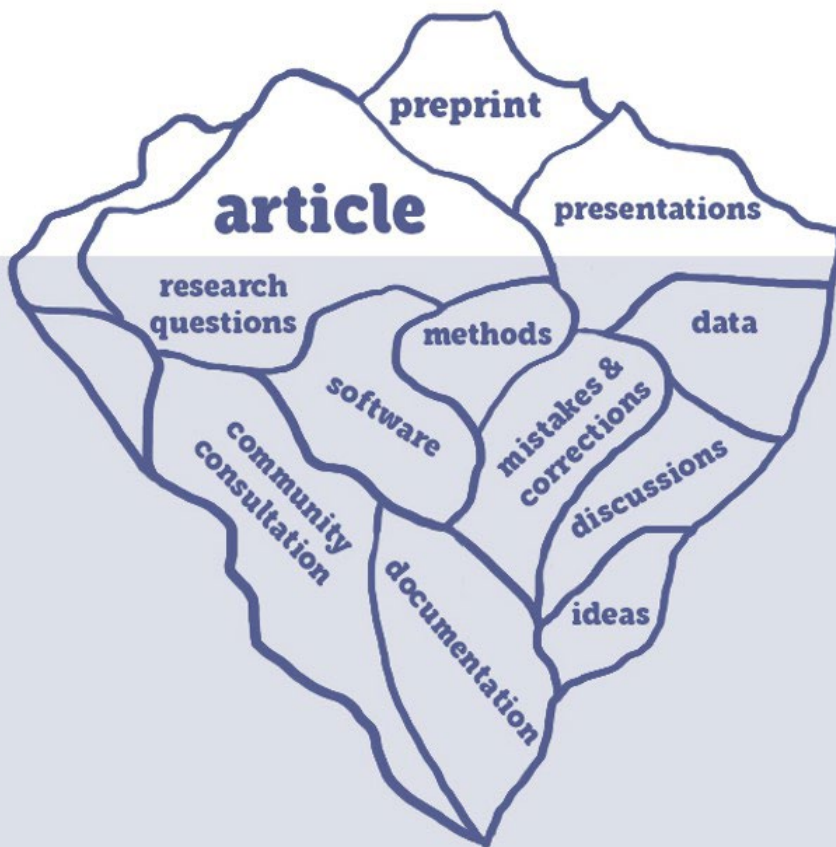
Long-term preservation
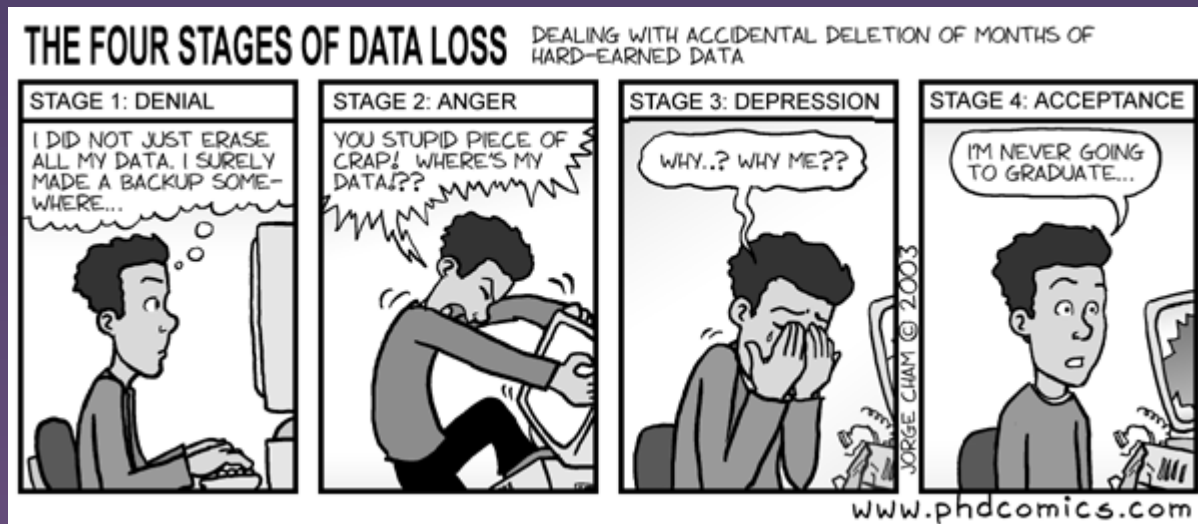
With **Open Data** the data repository is responsible for long term preservation and access!

The article is just a summary of research

Research Iceberg

Preventing fraud and inequitable research

Paleontologist accused of faking data in dino-killing asteroid paper

Scientist who oversees famed extinction-day site denies claims he made up data to scoop a former collaborator

6 DEC 2022 · 5:45 PM · BY MICHAEL PRICE

Bones of Black children killed in police bombing used in Ivy League anthropology course

How?

Open, Reproducible

What you've seen up to now:

protocols.io

Sharing of study methods/protocols on public platform

Data repositories

Slides available on Zenodo

(see slide 35 for an overview of data repositories)

Registered Reports

Peer Community In, Registered Reports

The Royal Society Open Science

Open Access

Via the journal or via preprint servers or institute repositories

SocArxiv

Peer Community in Archaeology

Licenses

FAIR principles

Slides/images are CC-BY

(see slide 36)

Community

Progress is not made alone

(slide 41)

Why not available upon request?

- Data availability decreases by 17% per year

- Chance of email address working decreases by 7% per year

'research data cannot be reliably preserved by individual researchers' - Vines et al. 2014

"**We received no response** to 41.3% of our data requests"
- Tedersoo et al. 2021

MadScientist
@MadS100tist

"Data will be available upon request"
Tweet vertalen

Meme explanation

33

Why not supplemental data?

https://estherplomp.github.io/2022-PRES-Data/#28

⊠ Data control: cannot be updated

⊕ Interoperability: not available in all formats which makes it difficult to integrate and interact with the data

☑ Availability: Difficult to access if the article is behind the paywall (supplemental materials are not included in the DOI and therefore the links can also break!)

🏆 Impact: Data should be a primary research output

🏺 Publisher requirements: Some publishers recommend using a data repository instead

Data Repositories

Archaeology Data Service (ADS)

Instructions for Depositors

Contact: Nicky Garland

Open Context

Archaeological data

IsoArcH

An Open Access and collaborative isotope database for bioarcheological samples

Zenodo

General repository for data, code, presentations, reports, training materials

Uploads up to 50 GB

Figshare

General repository

Uploads up to 20 GB

And more..

Find more using FAIRsharing or re3data

Licenses

Data

- [Overview](#) by Creative Commons

- [License chooser](#) by Creative Commons

[More information](#)

Software

- [Choose a license](#)

- [tl;dr Legal](#)

Creative Commons licenses by Foter (CC-BY-SA)

Data article

- Credit for generating/managing data

- Focus on methods and results

- References to datasets on a data repository (long term goal!), links to scripts and other outputs
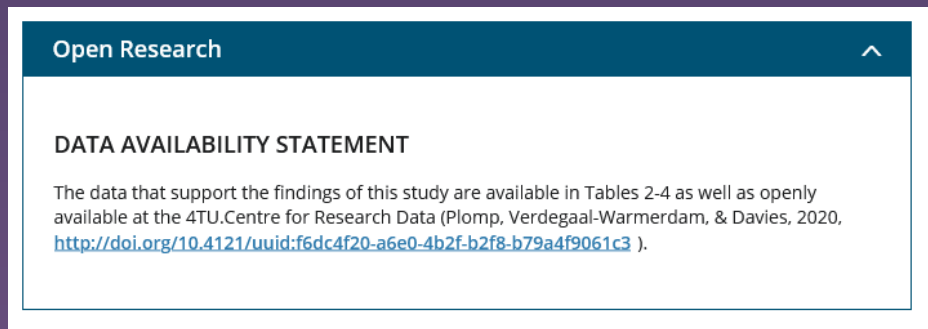
Data Article

Neodymium isotopes in modern human dental enamel: An exploratory dataset for human provenancing

Esther Plomp ✉ ⊕

Linking data/articles

Article: <u>Strontium, oxygen, and carbon isotope variation in modern human dental enamel</u>

Open Research ⌃

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in Tables 2-4 as well as openly available at the 4TU.Centre for Research Data (Plomp, Verdegaal-Warmerdam, & Davies, 2020, http://doi.org/10.4121/uuid:f6dc4f20-a6e0-4b2f-b2f8-b79a4f9061c3).

How to **link the underlying data and article**?

▫ Data availability statement and publishing the data and/or reserving a DOI. See <u>Linking Research Objects</u>

Removing Barriers to Reproducible Research in Archaeology

Peer Community In
Archaeology

RESEARCH ARTICLE

Open Access

Open Peer-Review

Removing Barriers to Reproducible Research in Archaeology

Emma Karoune[1,2,3*] and Esther Plomp[3,4]

- Introduce to reproducible research
- Examples of three levels of reproducibility
- Frequently Asked Questions

39

Research Data Management (RDM)

- Data Carpentry: **Data Organization in Spreadsheets** for Social Scientists
- Turing Way chapter on RDM
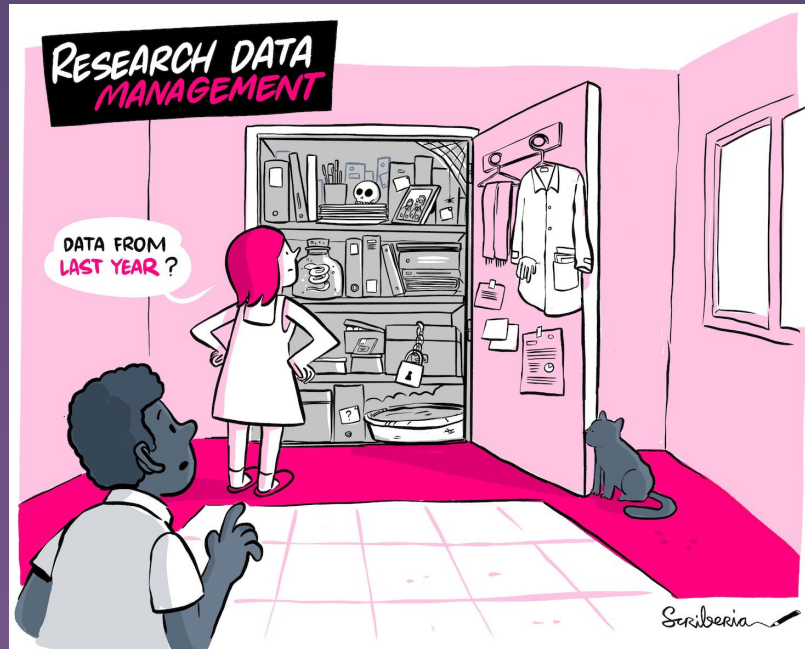
This illustration is created by Scriberia with The Turing Way community. Used under a CC-BY 4.0 licence. DOI: 10.5281/zenodo.3332807

Communities

Open Life Science

Applying open principles and becoming Open Science ambassadors

The Turing Way

A handbook and community for reproducible, ethical and collaborative data science

Anthro-data-science

An anthropology data science working group at UW

Rchaeology

Rchaeology is a community for archaeologists who use and want to learn R statistical software

- 16 week mentoring programme to learn all about Open Science principles and apply them practically to your research project

- Can range from publishing open access, to having educational platforms available to multiple partners

- Next round of applications likely opens in July.

- (PhD candidates from my faculty can follow this for credits!)

Intellectual Property, Indigenous Knowledges, and the Rise of Open Data in Australian Environmental Archaeology (Carly Monks)

Towards FAIRer phytolith data (Javier Ruiz Pérez, Juan José García-Granero, Carla Lancelotti, Marco Madella)
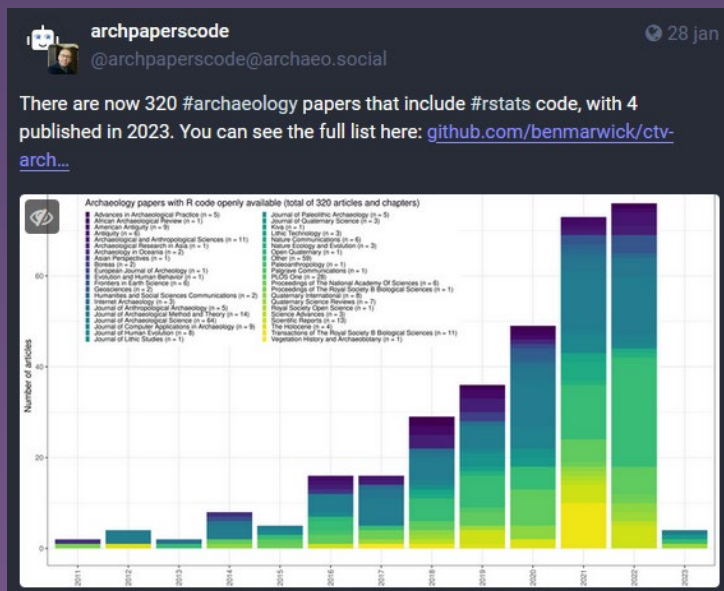
Succes Stories

Ben Marwick

▫ [rrtools: Tools for Writing Reproducible Research in R](#)

▫ Publications on: <u>Open Access</u>, <u>Data sharing</u>, <u>Replication</u>, <u>Research Compendium</u>, <u>R</u> and more

archpaperscode
@archpaperscode@archaeo.social   28 jan

There are now 320 #archaeology papers that include #rstats code, with 4 published in 2023. You can see the full list here: github.com/benmarwick/ctv-arch...

https://github.com/benmarwick/ctv-archaeology#publications-that-include-r-code

44

Ben Marwick

The lead author has some important recommendations to help us avoid this

Hi Ben,

Those certainly are interesting findings from that paper! I had thought there was something worthwhile happening with the sulphur results, but my p tests seemed to suggest otherwise. This kept me from getting support from co-authors for talking about these relationships between values as much as I would have liked. However, the exploratory graphs showed this interesting relationship. I think this ties into that idea that exploratory graphs are also really important and shouldn't be downplayed compared to p-values.

I wonder if I misread the Stata output (unlikely for such a low p-value) or if I, as you said, treated NAs differently in some way back then. This is also a good reason to put your stats code somewhere easy to find, as I don't think I really want to delve to look for it. Also a good reason to use something open-source like R, as even if I found it I wouldn't be able to re-test as I currently don't have a Stata license...

You're welcome to discuss these findings openly with my permission, especially since I don't think this reflects too poorly on me. I hope you find more folks making use of the APA system for reporting. My stats classes as an undergrad and master's student were in psych departments so I wonder if that's what influenced me.
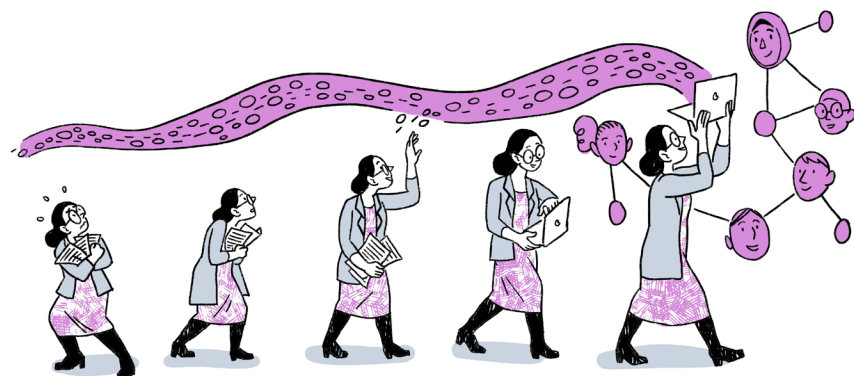
Best,
Chris

Chris Stantis

Open Phytoliths

website

- □ **Publishing guide**

- □ High **workshop/course attendance** (40-100 people) due to live translations

  - □ ICOPS openphytoliths YouTube channel
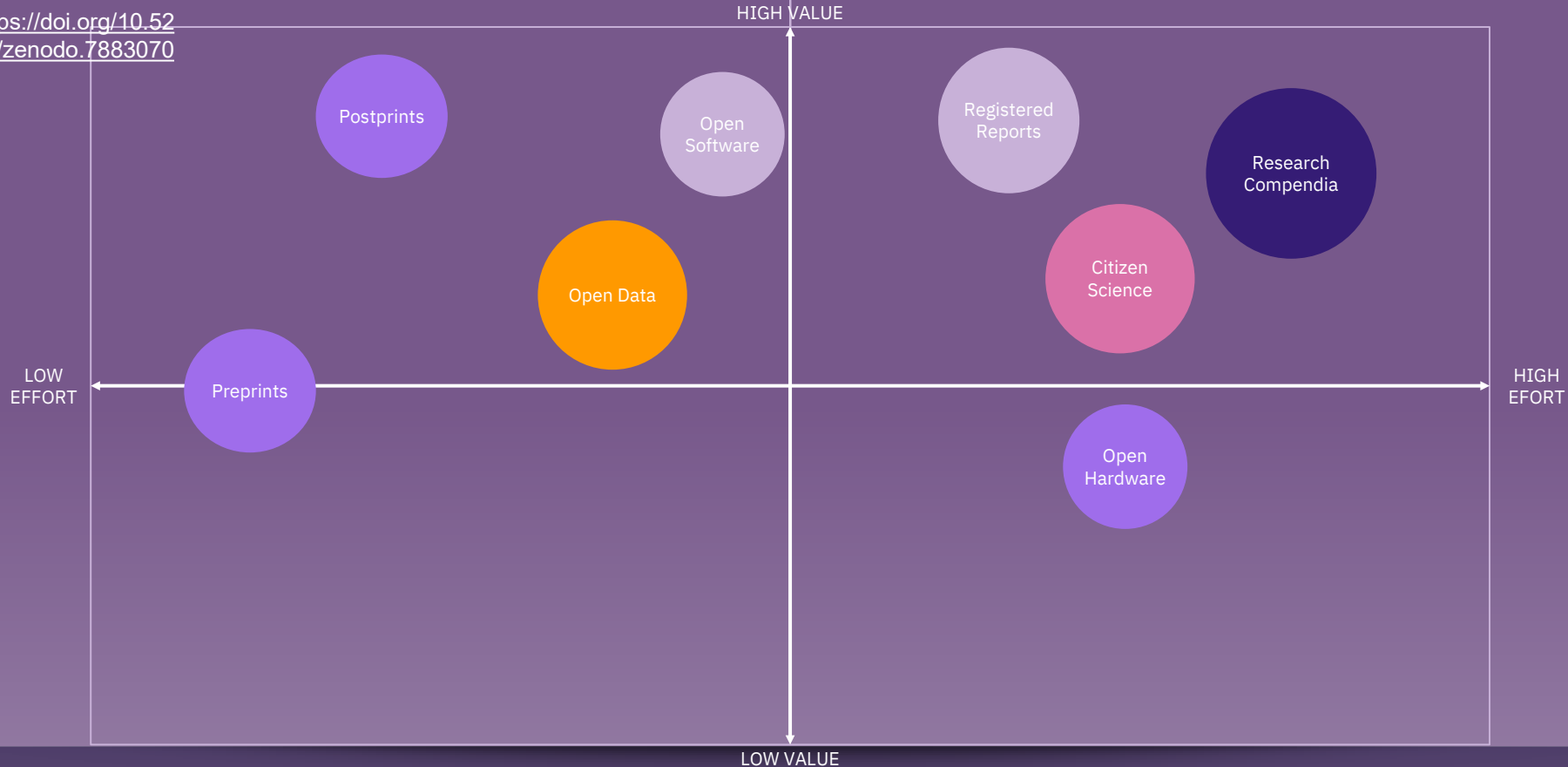
- □ **Article** that is open access + open data + code + methods

Journey

The idea is not to be 100% or fully open and to get there tomorrow

EVOLVING TOWARDS AN ERA OF OPEN RESEARCH

Scriberia

HIGH VALUE

Postprints

Open
Software

Registered
Reports

Research
Compendia

Open Data

Citizen
Science

LOW
EFFORT

Preprints

HIGH
EFORT

Open
Hardware

LOW VALUE

Starting where it makes sense

48

Thanks!

What questions do you have?

- @toothFAIRy
- e.plomp@tudelft.nl

Presentation template by SlidesCarnival

FAQ

Some more elaborated/documented answers to your questions:

https://doi.org/10.5281/zenodo.7320029

Appendix

Appendix A: Frequently asked questions about reproducible research in archaeology

You probably have many questions about different aspects of reproducible research. Therefore, we want to discuss the most frequently asked questions that we hear from archaeologists about reproducibility to try to remove barriers and help you make progress along your reproducible research journey. The easiest way to use this Appendix is to go to the question that is currently on your mind:

- How do you decide if you should publish data and/or code openly?
- Where do you start training skills in open science and reproducibility?
- What should you consider when you publish an article with underlying data/code?
- How do you share data to make it more accessible to others?
- How do you clean up the data and code before sharing this publicly?
- How do you share research methods more openly?
- What is metadata?
- What about licenses/copyright?
- Isn't reproducible archaeology more expensive?
- What if people misinterpret data or find a mistake?
- Is archaeology suitable for preregistration?
- A supervisor doesn't want to work reproducibility, how do you convince them?
- Will reproducible research be taken into account when looking for a next job?
- Do platforms like SciHub, ResearchGate, Academia.edu count as Open Access?