



MACQUARIE
University
SYDNEY · AUSTRALIA

Analysis before the spade

Brian Ballsun-Stanton

Faculty of Arts

Monday 17 November 2019

Analysis before the spade: Improving data quality by automating analysis before going into the field.

Book chapter is still with editors of *Digital Heritage & Archaeology in Practice*.
Contact me later if you want a pre-preprint copy.

Brief outline:

Some Philosophy

What could 'Analysis before the Spade' mean?

Some possible “easy” wins



MACQUARIE
University
SYDNEY · AUSTRALIA

Some Philosophy

I've found three philosophies of data so far [Ballsun-Stanton, 2010]:

- Data as communications, a container for meaning;
- Data as subjective observations, sense-impressions filtered by knowledge;
and
- Data as objective facts, measurements revealing the relationships of reality.

I'm *really* not sure which “Data” we are using today.

Prediction and Postdiction:

- Prediction describes a deductive, hypothesis-testing approach;
- Postdiction describes an inductive, hypothesis-generating approach.

If postdiction is conflated with prediction, it is prone to “fallibility of memory, motivated reasoning, and cognitive biases” [Nosek et al., 2018].

Deductive, Inductive, Abductive:

- Top-down (generalising *from* patterns),
- Bottom-up (generalising *to* patterns), and
- Rapid hysteresis between the two.

“Affordance is a property or feature of an object which presents a prompt on what can be done with this object” [Tubik Studio, 2018].

Philosophies of research inform our data's affordances and *what we think we can do with other folks' data*.

Marwick's Computational Archaeology [Marwick, 2017] could be a completely foreign land to someone used to a single (if long) report from paper-based single context recording.

If the data for analysis fits on your laptop, it's not big data.

We seldom can “run a script” and get tables, graphs, and rasters generated automatically.

We even more seldom have that script ready before we get to the field.



MACQUARIE
University
SYDNEY · AUSTRALIA

**What could
'Analysis before the
Spade' mean?**

Questions arising from field deployments of data systems to over 50 teams:

- How do we account for Archaeologists' time preferences?
- First mover disadvantage: how do we create network effects which make it desirable to use other peoples' data while in the field?
- What does the reproducibility crisis mean for Archaeology?



MACQUARIE
University
SYDNEY · AUSTRALIA

**Some possible
“easy” wins**

A program I've written has been continuously downloading social media data from October 24th. Before that program started its run:

- I knew the ending data structures
- I knew our desired outputs
- I had scripts written which could generate frequency tables, wordclouds, and plots.

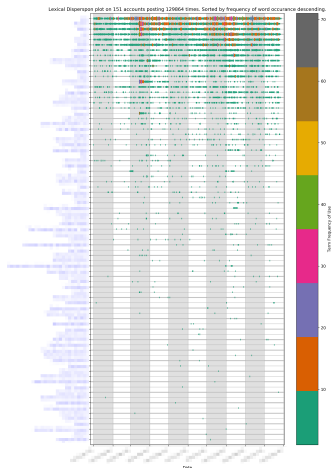
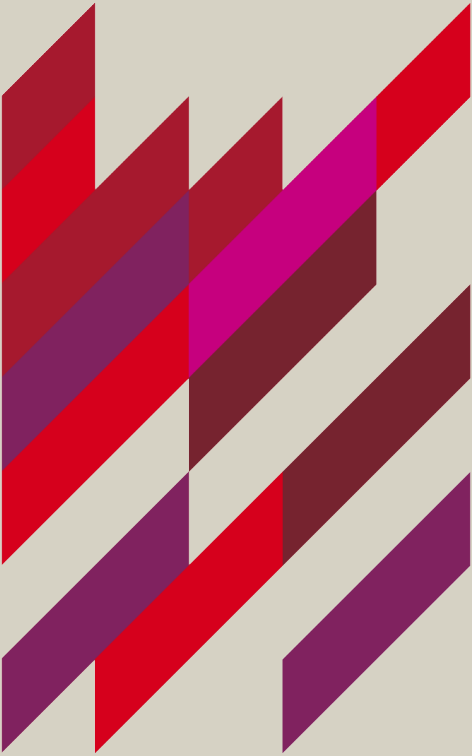


Figure 1: Lexical Dispersion plot by Brian Ballsun-Stanton. Blurred due to sensitive topics.

In order to take advantage of any data at all in the field, Archaeology needs to write much of its analytic pipelines: “Before the spade touches the ground.”

- Mandated reports are formulaic: generate the report is a press of a button.
- What reports do you wish you had nightly as a field director?
- How can a failure to replicate create useful controversy?
- How can your upcoming HDR students use your data in new ways? See: [Spring et al., 2018].



MACQUARIE
University
SYDNEY · AUSTRALIA

References



Ballsun-Stanton, B. (2010).

Asking about data: Experimental philosophy of information technology.

5th International Conference on Computer Sciences and Convergence Information Technology.



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.



Nosek, B. A., Ebersole, C. R., DeHaven, A. C., and Mellor, D. T. (2018).

The preregistration revolution.

Proceedings of the National Academy of Sciences of the United States of America, 115(11):2600–2606.



Spring, J., Maller, M., Mariconti, E., and Onaolapo, J. (2018).

An untapped resource to reproduce studies.

<https://www.benthamsgaze.org/2018/07/31/an-untapped-resource-to-reproduce-studies/>.

Accessed: 2018-11-27.



Tubik Studio (2018).

UX design glossary: How to use affordances in user interfaces.

<https://uxplanet.org/ux-design-glossary-how-to-use-affordances-in-user-interfaces-393c8e9686e4>.

Accessed: 2019-11-17.

Source code for this presentation is available at: <https://github.com/Denubis/CACHE-Presentation-Analysis-Before-the-Spade>

This work is licensed under a Creative Commons Attribution 4.0 International License.