Curating archaeological research data in practice, from the field to the archive

Zack Batist

INF 649: Digital Curation

November 4, 2024

zackbatist.info zacharybatist@gmail.com

About Me

Archaeologist:

- Used network analysis methods to investigate of prehistoric regional interaction
- Managed archaeological databases in fieldwork settings
- Maintainer of open-archaeo: a list of open source archaeological software
- Co-admin and community moderator for archaeo.social mastodon instance

About Me

Scholar of Scientific Practice

- Scientific data management practices
- Data-sharing as collaborative experience

Key Concerns

- 1) how open science reconfigures (or attempts to reconfigure) collaborative experiences;
- 2) the bureaucratic apparatus of open science;
- 3) the emergence of a new "informational" perspective among researchers, who are now made to think about their work from a curatorial lens; and
- 4) resistances against the power relations that emerge from these aforementioned developments

Curating Archaeological Knowledge in the Digital Continuum: from Practice to Infrastructure

Costis Dallas

From the journal Open Archaeology

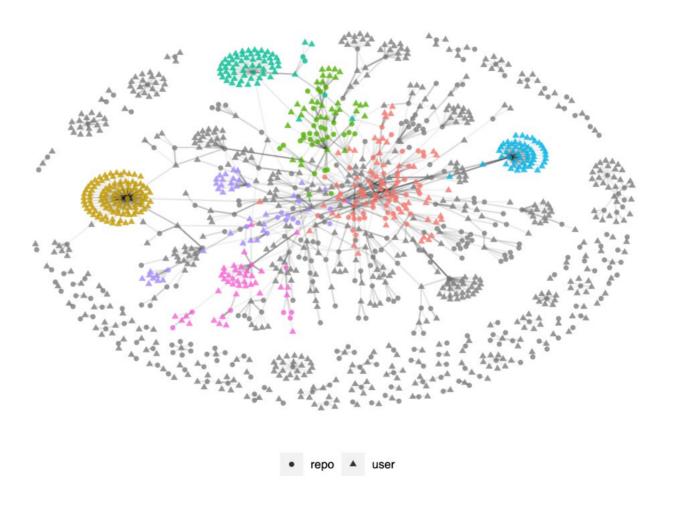
https://doi.org/10.1515/opar-2015-0011

- Framing data are connective tissue holding together a "continuum of practice"
- Research activities are acts of interpretation and re-presentation
- People who perform activities reconcile their actions with prior decisions, while anticipating future applications
- Activities are therefore loci of practical epistemic convergence
 - i.e. systems where meanings are negotiated in relation to communally-held objectives

Data re-use in practice

- Users lack sufficient context about circumstances of creation
- Context is often shared through informal or unrecorded media
 - e.g. at the pub between conference sessions, or via email
- Data re-use requires lots of additional work to make the data "fit" new questions and circumstances
- Not as simple as picking data off the shelf and running with it
- Social and professional norms influence <u>how</u> data are re-used, and <u>who</u> may re-use data

Tracing collaboration in open source research software development



Observational records

Conversational interviews

Retrospective interviews

Document analysis

Reflexive notes

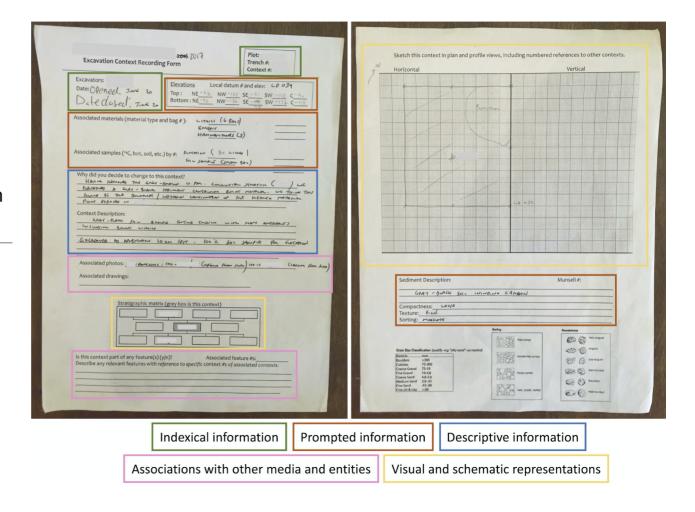


Data Collection

Recording Sheets

Theo: The context sheets and the trench report are more, sort of, stricter, in their [...] more defined in their, what they want out of you, whereas in the journal you can just write the fuck you want, pretty much

Lester: I mean I want to integrate with the project as much as possible, so the infrastructure that's in place is the stuff that I would use to interpret my data, so how the context sheets are laid out is how I record stuff.

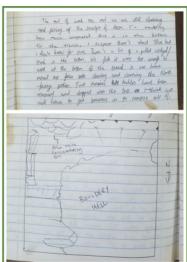


Field Journals

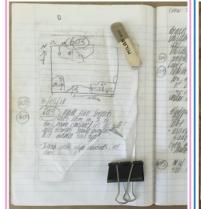
Theo: A journal is a stream of consciousness telling you where you're at and what you're doing in the trench and during the day. So some days it can be [unclear], other days when you're not actually doing very much it's just like half a page, a paragraph.

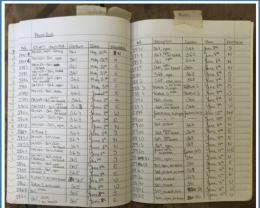
Zack: What do you mean by streams of consciousness? Just like what you're, writing what you're doing?

Theo: Yeah. And what you're thinking. So that you know, yeah, so that... It's because the idea of the journal is so that you can go back to it to understand what the excavator was thinking one [unclear] whilst they were excavating.









C

D

Comparison

• Ben: Umm generally I have tried to be uniform in the way I write my journal and my context sheets. I think the context sheets have to be a little bit more, I don't think—like not bare-bones, but like less descriptive than the journal. So, for example, like the sediment descriptions, just, it's gonna have like one word for each of the categories that Alfred wants. Uhh where in the journal I can provide more detail, and if people want to look back at it they can.

• Liz: [Describing recording sheets] So like for like the labelling of units, I give them something that's very very vague, that purely describes what we're seeing, rather than interpreting what it means, so just soft fill on that side of the trench, for example. And then in the actual, in my trench notebook, that's where I put all of those [interpretive] thoughts in more detail.

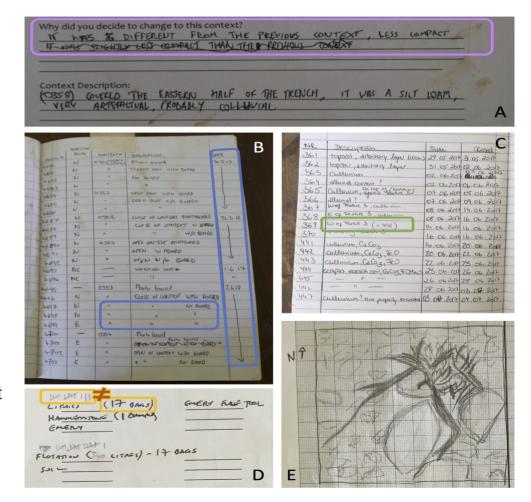
Digital Transformations

Databases

- Centralize all structured information generated by the project
- Enforce consistent data structures
- Scaffold the implementation of rigorous methodological procedures
- Synchronize the flow of archaeological materials and records across work settings
- Ensure that data maintain their value in perpetuity
- Custom built, backed by improvised skills learned on the job

Translating to the DB

- Data entry and data cleaning translate from paper recording sheets to homologous tables
- Must balance rough fieldwork conditions with smooth computational interfaces
- But field records are often inconsistent and messy
- Data entry inevitably involves modification, omission and data loss
- This produces anxiety in data entry workers, who must make editorial decisions they are not prepared for



Jamie: ... on a complicated project like this one, or just like any kind of real work field study, it's not gonna be ideal for data collection. Like people are gonna record things differently, or improperly, and it's gonna have little, kind of uhh, and it's a long term study like this one, things are going to change over time and it's going to be inconsistent. So in that sense, <u>fieldwork, or field data is never going to be ideal,</u> perfect, clean for data management and data analysis. But umm, ways to get around that, I think, are implementing, as early as possible, or, you know, as consistently as possible, some kind of standard. So like, we have the entry form, so it's telling them, like, what to say, where to say. We have the drop down menus so they, they're not just going to write something random, they're gonna have a controlled, kind of vocab, and that makes it easier, I think, for everyone to interact with these digital tools, if it's been laid out for them.

Zack: Sort of like prompts.

Jamie: Yeah exactly. So it's not just here's an open ended thing, write a page. It's like fill in this specific information in this way.

Zack: Why is that necessary?

Jamie: I mean, just in terms of collecting useable information, right. If someone, like I know when I want to analyze data, it has to be the same time. So something as trivial as like capitalization, or if someone writes a sentence on something and someone else is reporting it as like initials, you know, it needs to be in the same format so you can actually work with it.

Zack: It's a certain understanding and way of working that-

Jamie: Exactly.

Zack: Would you say that this overcomes a certain disconnect?

Jamie: Yeah, it's tough, because like, ultimately, the data analysis that we're doing is, it's for the archaeology. **[The fieldworkers] are not doing archaeology for, you know, data purposes. But I think ideally they would be a little bit more connected, and stuff could be recorded, like keeping data in mind, and keeping the fact that you have to work with this stuff later, in mind.**

Digital Archives

- Delegation of data cleaning, curation, preservation and distribution tasks to archival services
- Project directors were indifferent about data re-use
- Priority was to enable archaeologists to move forward with new projects

- Effective re-use will always require discursive engagement
- Focusing on the technical processes of data curation is a way of avoiding this concern

Take-Aways

- Projects implement digital and analog tools to reduce subjective representations
- However, subjective and informal communication persists as fundamental aspects of archaeological research
- Formal representations are lossy simplifications
- Communication is not transaction, and involves parsing context and subtext based on shared experiences
- There is no universal audience; archaeological repositories have implicit designated communities in mind

Thank You!

Questions?