

Reflection on CIVIC EPISTEMOLOGIES

Breandán Knowlton, Historypin

20 Feb, 2015

CITIZEN SCIENCE

- Uses the concept of “Citizen Science”
- We usually use the idea of “citizen historian”, but the meaning is the really the same in this context.

Historypin fits with

- Contextualisation: citizens submit data such as letters, stories, films or photographs to gather a meaningful context
- Complementing collection: citizens submit data to complete them or make the collection grow. Do this locally.
- Classification: citizens tag the data or label it to group similar material and make information more easily retrievable in the future.
- Co curation. Mostly with the aesthetic arts, but in historypin this involves choosing to put photographs into sequences which can be shared as a unit with other users. This is one of the most popular features on historypin, 3x as popular as simple text comments.

We think of geotagging as a special case of this. Some thinkers in the world of digital cultural heritage have encouraged collection holders to get “beyond the search box”.

Maps create a single pivot point that transcend questions of tagged vocabulary or language, and do not rely on preconfiguring a list of questions for users. They let us explore at every point, jumping off to new directions and potentially connecting with other citizen historians.

— HOWEVER: In an important way what we’re doing isn’t science at all, as we don’t always lead our projects with a professional researcher. Sometimes we do (Stanford, Yale), sometimes we are led more by a collection holder or a community group.

Citizen Scientists described their association as being with “activist organisations” - genealogical societies in the case of people interviewed. There is a Confusion between citizen science and crowdsourcing — this is a theme.

We would suggest that this is actually a continuum - for example, WikiProjects are often organised around a research topic with a professional researcher. We REJECT the idea that meaningful research can only be done with a professional lead.

ACCESSIBILITY

You’ve indicated that the Collections side of things should be equally available - that this isn’t the exclusive domain of Archivists. The same probably holds for researchers.

Our experience is that successful projects in this space require four things, each active:

Community (of citizens)

Collections (of heritage)

Context (often academic or research-led, but not necessarily)

Cash (a funder with an interest in the outcomes of the project)

The Roadmap shares some necessary preconditions:

- results have to be open
- technical facilities have to be in place
- planning has to be done in cooperation with citizen's groups

We completely validate those Roadmap findings on accessibility.

The Roadmap vision seems to be quite communitarian and to represent a partnership model of connection to citizen scientists. I would note that SMEs and cultural organisations are not necessarily technically unsophisticated — but they may use tools that are quite different from the larger research infrastructures used by universities.

To answer the question: “how can Humanities-based research in which the citizen is invited to play an active role, support re-conceptualization of the ways in which cultural heritage can reflect, construct and enrich individual and collective identities?” This is a tough one! But the first step is to recognise that local identities are created through the lens of local goals, and supported through a vibrant and associative civic life.

Anything that can be done to reach out to and strengthen associational life can help to build the bridges that we need to heritage, digital and otherwise.

If culture needs to be accessible to researchers, then research and the practice of it needs to be accessible to cultural actors, which is all of us. Research is not just the domain of the professional - there is a continuum of volunteer and professionally-validated practice.

INFRASTRUCTURE

The Roadmap idea of common layers for trust, support services, persistence seem ambitious and perhaps don't match the reality of collections and platforms, at least for small and intermediate platforms like ours.

Taking a cue from the Architecture Development Method, the roadmap emphasises the need for common platforms. We disagree. The nature of the internet, our largest tool for collaboration in the history of humanity, does not tend to do well with centralised infrastructure. Rather than trying to harmonise requirements, we suggest thinking about lowest-common-denominator interoperability. Silos are actually fine, as long as you can get goods in and out of them.

TRUST

Trust is a huge issue here. There's no reason to think that my small NGO would want to trust a key piece of our organisations's mission to infrastructure that's being run by someone else, even for well-intentioned reasons. Our experience with academic projects that come and go have convinced us that it is far safer for our objectives for us to buy cloud

computing and storage services from commercial providers, at commercial rates, and simply try to offset these costs using a combination of earned income, unrestricted grant funding and research & development restricted grants.

A truly federated service model that you have proposed is not something that our organisation would be able to participate in fully.

However, if you can replace a single or multiple elements of our contracted infrastructure services with something that does essentially the same thing in the same way, we can start to consider the longevity of the solution and the nature of the partnership.

I thought it might be useful to outline the nature of some of the e-Infrastructure elements that we use.

- We have cloud-based computing services provided by Google, cloud-based structured data services provided by Google, cloud-based index servers running on the open source ElasticSearch platform and hosted by a third party, Digital Ocean, and a third-party map-tile-serving service provided by Klokan Technologies.
- We support Klokan for tile serving rather than Amazon S3 or one of the Google options because Klokan is a leader in the space and are doing useful projects to unlock the value of maps to heritage research, so we're happy to pay them for services in part to support the mission. But this emphasises the need for trust.
- I trust Google for some services because they're a huge company and they have an economic interest in keeping clients like me happy. Even as a small NGO, I'm free-riding on the set of services and level of support that Google has built out to support enterprise clients. I use Klokan for tiles because of who they are, specifically Dr. Petr Priedal, the founder. I know of his work with lots of collections, and his combination of for-profit and non-profit work. That makes me trust his infrastructure.

If you want to tempt people to change bits of their infrastructure, including to federate or distribute costs, you'll need the drop-in replacement to do the same thing in the same way. For example, the preservation API for the Internet Archive uses practically the same method signature as the S3 cloud storage service from Amazon. This is a good candidate for replacement — the infrastructure does the same thing in the same way, but with an organisation that has sufficient long-term foundation funding to show sustainability, and an ethos that is more closely aligned with our organisation.

Incidentally, you should note that the Internet Archive is prepared to host almost unlimited amounts of material, for free, for anyone with a collection that meets their terms of use and is generally accessible. They'll even host material for archiving purposes that's not publicly

accessible if there are rights issues with the content. Just worth bearing in mind, since the Internet Archive has a head start in this area over eInfrastructures that might be developed in Europe.

The trust model for any new platform has to extend further than just to how you perform user authentication. The trust model has to include why the collection holder or service provider would want to trust you with mission-critical data in a way that is cost-effective for them to execute and monitor.

USER EXPERIENCE

And one small note on user experience for end-user facing platforms. Getting the user experience right for contributions by a knowledge community is incredibly difficult. It requires endless research, testing and tweaking. The version of the First World War project tool that we've deployed still has lots of usability issues, but is the third complete rewrite of the system from the ground up in the space of a year. Each time we redesign it in reaction to user feedback the platform improves, tests better and gets more user engagement traction in analytics, but this is a very detailed and long-running process just to solicit expert contributions to one very narrow domain. The resulting product is usable, but perhaps only in this narrow context. Trying to create common user experiences for seemingly-simple tasks like tagging or geolocation that are widely applicable run the risk of pleasing no-one. The recommended model might do well to consider that topical domains will probably still need their own, very specific interfaces, and that it might be more useful to invest in common storage and processing infrastructure than to try and support a common web-based client for a variety of tasks in different knowledge domains.

CONCLUSION

From the document, "This report raises more questions than it can answer." Of course, it's an early stage for this report.

But with an open-minded discussion, and the participation of the kinds of experts who are in the room and participating remotely, I'm confident that we can find sufficient common ground to advance digital humanities research in a way that respects the unique contributions of each member of the community.