PENTANDRA

DRAFT PROPOSAL [2017-05-01]

Scholarly commoning prototype Proposal

2017-03-27

https://pentandra.com/solutions/scholarly-commoning-prototype/

DRAFT PROPOSAL [2017-05-01]

CONTENTS

1 WHAT IS NEEDED FOR A SCHOLARLY COMMONS?

> If I had asked people what they wanted, they would have said faster [preprint servers].

– adapted from a quote misattributed to Henry Ford, who improved the process of automobile production (workflow) such that the automobile became a practical means of conveyance for many more people. The Scholarly Commons is a culture, but we're missing critical infrastructure to tie everything together and allow the culture of the Scholarly Commons to thrive. It is not enough for disciplines to 'spin up their own variety of commoning' (Bruce Caron). In order to tie everything together and avoid perpetuating cultural and disciplinary silos, we need 'technical duct tape' (also Bruce) harnessing the power of the Web to provide the machine interoperability needed to tie together disciplines, cultures, languages, questions, answers, inputs, outputs, data, sources of data, evidences, standards, principles, practices, tools, projects, organizations, funders, individuals, interfaces, and expectations across the globe in an open and unified way.

The Scholarly Commons should be a jumping-off point to new ways of working together, taking the best ideas of how to do scholarship in a modern setting, setting aside as many of the undesirable social, cultural, and technical limitations as possible. According to Diffusion of Innovations we need a viable alternative that can be trialed, and I'm not convinced that we have an existing viable alternative, hypothetical or otherwise. To this end I am proposing here a viable alternative, distinct and comprehensive enough to be considered an alternative system of scholarly communication. I am proposing to create a prototypical research system in the spirit of Concept II of the Doable Pathways to the Scholarly Commons.

Decision trees are a good start, but they are not enough on their own to create a viable alternative to the current system. This alternative system must be better than anything currently available, so there is real motivation to leave behind old practices and adopt new. The theme of this prototype will be the Scholarly Commons, with an emphasis on creating generic, domain-agnostic

components, so that the ideas and functionality created here can be applied to as many domains as possible. The work is composed of three interrelated projects:

Open-research curation system

Question-driven approach that exposes and preserves the researcher's process

Research guides

Semi-automated decision trees that guide and assist research

Branding and visualization

An integrative approach to the design and functionality of `scholarlycommons.org`

Cost: TBD

2 OPEN-RESEARCH CURATION SYSTEM

See Defining the Commons Platform and my 400 word response for some background thoughts, though this is still a moving target.

A vision for this project is painted beautifully by the humanist Jeffrey Schnapp in his concept of ubiquitous curation in the lecture Knowledge Design.

The most fundamental purpose of this project is to provide a place to explore and eventually answer the question: *what is the Scholarly Commons?* It will build upon all the work that we have done in the past as a steering committee and working group, as well as the outputs from the workshops in Madrid, Portland, and San Diego. Going forward, it will provide an open and even playing field for anyone to contribute to the definition of the Scholarly Commons.

The output of this project will be a modern, reusable, prototypical research system, aligned with the culture of the Scholarly Commons, that could be used for accomplishing Concept II (focused on the research process) of Doable pathways to the Scholarly Commons, discussed during the San Diego workshop. The Scholarly Commons should be a jumping-off point to new ways of working together, taking the best ideas of how to do scholarship in a modern setting, setting aside as many of the undesirable social and cultural limitations as possible.

One of my greatest takeaways from the main track of the San Diego workshop was that because the research processes were made explicit and studied this way, it allowed the workshop attendees to reason about, find patterns, and learn from them. With this end in mind, a central feature of this system will be exposing the researcher's path explicitly. From the researcher's perspective, the entire research process could be viewed as a sequence of decisions. These decisions and the process of how they came about should form the basis of research publication and collaboration. It is the *process* that needs to be open and replicable, not the tools (as much).

If we don't make that process explicit, how will we be able to share, reproduce, or assess the differences? If we keep the process implicit, or even leave it to post facto analysis to attempt to reveal it all alone, we are missing the whole point, and we end up not very far from where we are now: publishing *about* the research, instead of publishing *research*, and our ability to effectively common will be impeded. What is needed is an infrastructure where, as Jeffrey Schnapp says in his presentation on Knowledge Design, the process *is* the product.

The general approach will build on the early work of on Issue-Based Information Systems (IBIS), Knowledge Media Institute's Scholonto later ideas, other thoughts from the brightest minds on the future of scholarly communication, and our own research on how to fit this all together that we have conducted over the past several years. It will implement the best thinking on evidence and argumentation in a collaborative research context, bringing in bits of knowledge from around the Web, including existing declarations (building on WP1), and mapping that knowledge together (building on WP2), all within a beautiful, simple, and intuitive experience.

Wherever possible, this system will integrate with and connect current tools, standards, and implementations instead of replacing existing functionality.

2.1 Tentative high-level components

- Persistent identifiers for:
  - Researchers (authentication via ORCiD and IndieAuth, or another Web-centric equivalent, for the indie people)
  - Organizations and groups (via [Grid])
  - Concepts
  - Questions
  - Everything else uniquely identifiable :-)
- Design an ontology, tentatively called the Research Intent Ontology (`rio`), that describes the process of knowledge design in research. It will be loosely based on ideas from DecisionML, IBIS, Scholonto, and the recent Design Intent Ontology

- Develop a user experience around the flow of research on the Scholarly Commons, including questions, peer-to-peer interactions, and versioning, that will form the basis of discourse as well as inform the development of the above ontology
- Create integrations with common tools, starting with the tools that we used during the course of this program, such as Google Docs, Hypothes.is, [Slack], and [Trello]
- stuff would happen via Hydra
- Use A/B testing to compare processes, ideas, and user interactions to find what works best

2.2 The fundamental building block of scholarly commoning

The basic premise is that our current outputs, implementations, and approaches to scholarly communications, such as papers and working groups, are inherently closed, and not in harmony with the culture of the Scholarly Commons. Openness is more than just licensing and filetype compatibility, or even open admission to a knowledge club. We need a way to open the process so that people can get involved and can collaborate without the barrier of time or place.

The question is the most fundamental building block of scholarly communication. All researchers question. Questions are the bedrock upon which all answers are pursued. Discovery happens after finding the right question to ask. Making public one's own questions exposes the mind of the researcher, and is the mental and cultural equivalent of a teacher flipping a class. Questions partition research into tractable units of work. Questions are the intellectual undercurrent of the research process. Questions invite the culture of openness that we are so desiring to cultivate.

Figure 2.1 A researcher's stream of questions

This is not just another application. It is not a Facebook for science. I think this approach is vital to the integrity of the Scholarly Commons. We need to expose research at the cellular level. We need to expose it at this level of granularity so that other people can involve themselves freely in research. Otherwise we're still just dealing with closed groups, artifacts, and processes.

2.3 An organizational microstructure for research

The Research Object approach is fine for publishing a structural depiction of research data, but what we need is a conceptual understanding of what the researcher is trying to accomplish.

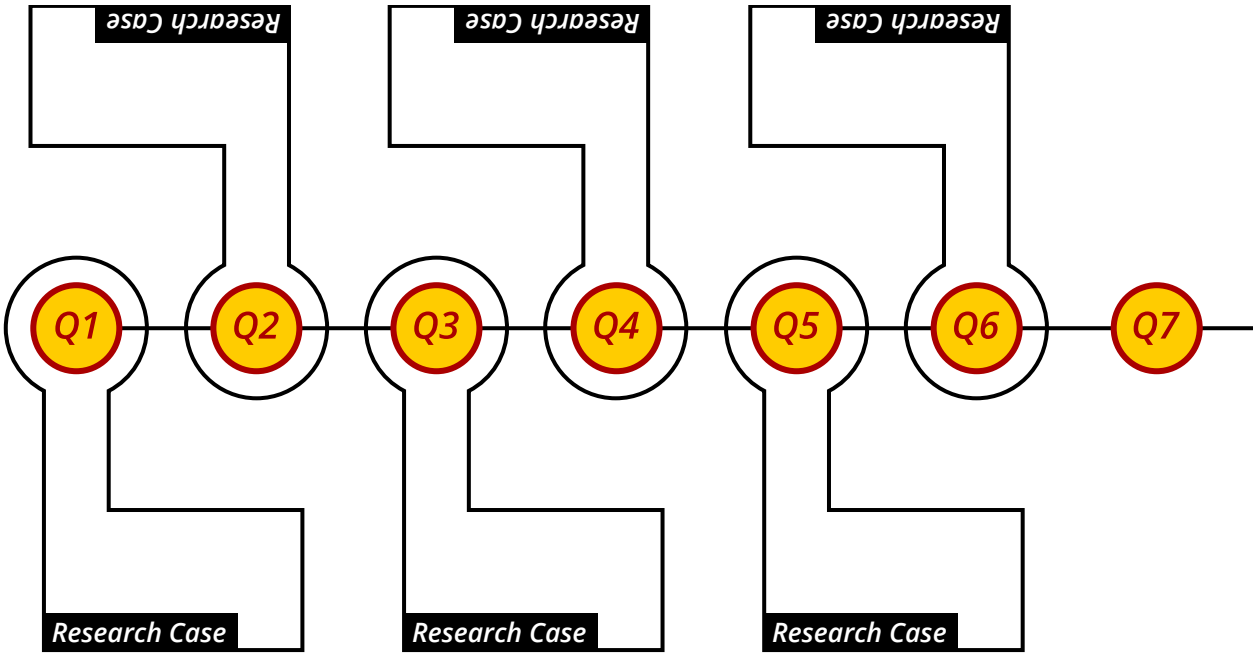

Figure 2.2 Questions with research cases

Cameron's TCP/IP for research.

3 RESEARCH GUIDES

See Infrastructure Map Proposal for some background thoughts, though those ideas have evolved into something much more simple and distributed, thanks to the discussion on decision trees that came out of the San Diego workshop.

Decision trees are action-based, describing patterns of behavior, not patterns of logic.

- the process trees would be a scholarly output of their own (WP3)
- each workflow would be exposed as a community best practice for a specific situation
- and also as the actual 'what happened' for a researcher
- these two would form a processual backbone of the Scholarly Commons
- and would make research reproducible

We've started by selecting the entities of data and software tools, but tracking decisions could be useful during other parts of the research process.

myExperiment and Taverna

Explicate the development of every type of research output

Research guides. At any point, you should be able to see what the plan was (or that there was no plan), and how plan that differed from what actually happened.

Semi-automating

Reusable workflows

Research zealots will still find ways to blaze trails where there are none, but why not plan for the city and prepare for the streets to make it easier for everyone to transition? We need streets for machines for navigate, so people don't have to. The tools that are part of the commons should be

able to discover other tools and people and research with which to interact. Our tools and machines need to know how to common too! Why not? All this could transform research into a realtime dynamic system rather than the current static system against which we struggle.

3.1 Human-curated decision trees

3.2 Machine-curated patterns

As more decision trees are curated and research done, use machine learning to find patterns of commoning and then generate decision trees based on these patterns, with the goal of assisting future research or decision tree design

4 BRANDING AND VISUALIZATION

See Branding the Scholarly Commons *for some background thoughts, though those thoughts have evolved since then somewhat. This is the only part of this proposal unique to* `scholarlycommons.org`.

As with all components of this project, we are taking a two-pronged focus with `scholarlycommons.org`: domain researchers, and the public.

Logo

4.1 Visualization

See Infrastructure Map Proposal *for some background thoughts on this, as well as some* earlier emails.

The visualization builds upon the data and understandings obtained from the open-research curation system and the research guides to help people (and machines) see how the commons works, or how the resources of the commons interact. Though scholarly commoning will be happening across all disciplines, in some form or another, `scholarlycommons.org` will be the place that we study *how* and *why* this commoning process actually happens, across all domains. Think of the people studying the scholarly commons as a collective.

The visualization is made of representations of the types of resources used in scholarly commoning, both interactional and artifactual. These resources are represented visually as smaller components of a greater whole. The components are a visual typology that will aid in understanding not only the details of the research process, but how the culture of the Scholarly Commons provides a more fluent and integrative research experience. Pieces of the research process only fit together in a certain sequence. This visual typol-

ogy will depict and reinforce the sequential nature of how research actually happens or could potentially happen.

The overall visualization is dynamic, as new types of resources are discovered, and is based on an underlying map of the resources accessible to machines. It will connect things that have previously been unconnected. At any given point in time the map gives a picture of what the commons is at that particular point in time. It is not a top-down rendering of what we would like it to be, but a versioned representation of the actual resources and potential interactions between those resources. At first these resources will be more domain-agnostic, but over time, more and more domain-specific resources will be represented.

5 OTHER FUNCTIONALITIES

- Web pages
- Blog

6 GOVERNANCE

This project is to be undertaken as an ongoing partnership between FORCE11 and Pentandra, rather than as a one-time outsourcing to a contractor or vendor.

Rather than assembling an ad hoc advisory board or organizing a working group, invitations to collaborate directly will be offered. Discussions and research as to how the system should work should happen within the system. In this way, we hope to receive better and quicker feedback, and in the end develop something that actually works well, and builds on the ideas of many people directly.

7 MAINTENANCE

8 ACCESSIBILITY

9 SELF-SUSTAINABILITY

This system will serve as a testbed for new forms of research funding. By exposing more of the research process, funding can happen at a smaller granularity, lowering the risk for funders while at the same time offloading some of the administrative overhead of the funding process. We see a lot of benefit in heading in this direction with research funding, especially if it provided the flexibility to accommodate specific disciplinary and cultural approaches.

Besides acting as a market for research funding, another possibility that could be explored is to provide contextually specific opportunities to connect, for example, service providers to researchers.

If we can find patterns of sustainability that work well and are commons-compliant, these patterns could provide valuable revenue streams that could sustain not only `scholarlycommons.org`, but could be replicated as others build similar collective-focused, self-sustaining systems to answer other specific questions, and could serve as a scalable means to fund ongoing research and development of the Scholarly Commons.

10 A TENTATIVE DEVELOPMENT SCHEDULE

The work will be organized into 11 two-week sprints. All three projects will be developed incrementally and simultaneously, as they will inform each other during the development process.

- Sprint 1: Feb 13–24
- Sprint 2: Feb 27–Mar 10
- Sprint 3: Mar 6–Mar 17
- Sprint 4: Mar 20–Mar 31

(Spring Break: Apr 3–Apr 7)

- Sprint 5: Apr 10–Apr 21 (Everything desired for the cc Summit must be completed by the end of this sprint)
- Sprint 6: Apr 24–May 5
- Sprint 7: May 8–May 19
- Sprint 8: May 22–Jun 2
- Sprint 9: Jun 5–Jun 16

(Summer Vacation at Yellowstone: Jul 3–Jul 7)

- Sprint 10: Jun 19–Jun 30
- Sprint 11: Jul 10–Jul 21

A ABBREVIATIONS

CC	Creative Commons
IBIS	Issue-Based Information System
IP	Internet Protocol
TBD	To Be Determined
TCP	Transmission Control Protocol