"Invisible Work" as a Lens for Understanding Humanware's Role in Research Cloud Computing

Evidence from an interview-based study

Dan Sholler, PhD

Project Scientist, Technology Management Program

College of Engineering

University of California, Santa Barbara

dsholler@ucsb.edu

My research interests

I systematically interview, observe, and survey people who are adopting and using new(ish) technologies, with the goal of documenting pain points and unintended consequences. I try to suggest some ways forward for workers and organizations

- Records managers in city government; doctors, nurses, and other healthcare professionals using EMR; most recently, scientists who use and develop open source software and cloud computing
- I aim my publications at fields such as Management and Organization Science, Information Science, CSCW, Social Studies of Science and Technology

"Who's doing the dishes?"

The process, as I see it

Public cloud vendors offer products that help researchers scale, innovate, or otherwise improve research processes

⬇

Vendors learn from research applications and associated challenges, presumably build better tools and services
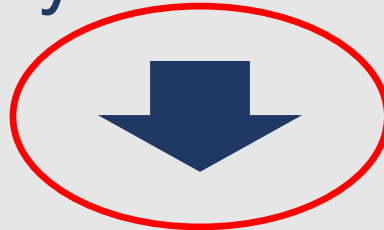
⬇

Vendors profit

The process, as I see it

Public cloud vendors offer products that help researchers scale, innovate, or otherwise improve research processes

⬇ ?

Vendors learn from research applications and associated challenges, presumably build better tools and services

⬇ ?

?

Vendors profit

"Who's doing the dishes?"

- Researchers, too, benefit immensely from cloud computing offerings
- But it is difficult to quantify the research value of any technology in science, though folks have tried:
  - 'NSF made 18,592 awards totaling $9.6 billion (1995-2016) that topically reference "software" in their abstracts.' (Katz et al., 2018)
  - Various surveys support the idea that many researchers simply could not get their work done without a given technology (e.g., the Software Sustainability Institute surveys)
- And perhaps more difficult to assess the costs of meshing cloud services with existing research workflows and collaboration patterns

What is invisible work?

No work is inherently either visible or invisible. We always "see" work through a selection of indicators: straining muscles, finished artifacts, a changed state of affairs. The indicators change with context, and that context becomes a negotiation about the relationship between visible and invisible work. With shifts in industrial practice these negotiations require longer chains of inference and representation, and may become solely abstract.

- Susan Leigh Star and Anselm Strauss, 1999

The central argument of this presentation is that much of the work required to migrate and sustain cloud research is rendered *invisible*, and that the work/labor is not currently shared appropriately among stakeholders. Researchers do the dishes, for now; vendors should do more, for both normative and economic reasons.

Outline of the study

- Guiding questions, at the start:
    - What are the barriers and facilitators of adopting cloud resources for research?
    - What are the pain points of adopting cloud resources?
    - How do researchers and research support staff overcome the barriers to cloud resource adoption?
- Refined research questions:
    - What types of "invisible" labor enable research cloud computing?
    - How do the workers who perform "invisible" labor perceive the work and its potential impact on their careers?
    - How is this work fed back into vendor learning processes and actions?
- Interview-based study of researchers using cloud computing in one discipline at one R-1 university (total of 45 interviews)
    - Brief histories of the research pre- and post-cloud usage
    - Critical incident questions
- Qualitative data analysis, grounded theory approach

Two types of invisible work

1. Managing billing for multiple projects

2. Migrating to the cloud

Managing billing for multiple projects

- Billing data sent in batches that are not clearly itemized
- Requires some Python code to parse the billing information and translate it into university billing structures
- This work is often not written into grants
  - Unclear who is responsible in the first place
  - Furthermore, it limits the accessibility of the tools to well-funded, well-supported researchers

Managing billing for multiple projects

"So example, over the first three billing cycles, *I wrote tens of thousands of lines of Python code to parse the billing documents* [public cloud vendor] *sent us.* It's getting a little better now that I've got it down, but I'm still revising that code, and *I don't have any time to document it.* So if someone takes over my job, they're going to look at this and probably want to kill me. But it's the reality. They send us back stuff that, manually, I'd probably never be able to say "Professor X's lab used this amount, and this is what comes out of their budget." I'd just have to charge everyone the same, which is maybe looking like it will be the case down the line when vendors get more exclusive access to our universities."

Managing billing for multiple projects

*"That work, **0 hours were written into any grant**, it's not written into my job description here* [at the support lab]*. And I'm not getting a break on anything else I was doing. So we really, we are all-systems-go for the computational part, the research part, but we are no-systems-go for the administrative stuff."*

Who does this work fall on?

- Graduate students

- Postdocs

- Distributed types of IT staff

- Staff researchers

- PIs

- HR

- …

Migrating to the cloud

- Moving years or decades of work to the cloud is not easy [he says in the direction of the choir]
- PIs in exploratory interviews said they'd need to hire a postdoc or graduate student full-time to migrate
  - None actually hired a dedicated person
  - But postdocs and graduate students were doing that work, on top of their expected work products

Migrating to the cloud

"*Had I known I was going to take on all of this admin work, like figuring out for my PI how much data storage we will need years beyond when I'll be here, I would have kept my mouth shut* [about being able to do this work]. **It's a headache and it keeps me from doing things I'll actually get credit for**." [postdoc]

Migrating to the cloud

*"I'm a, how is it called, **a guinea pig or a mouse for the future researchers who will do everything in the cloud**. I do think one day it won't be so difficult to port everything, or it will just be done there from the very beginning. Now, it's a lot of work."* [graduate student]

Some insights for researchers

- Many of the things I heard from interviewees echoed what I've heard from scientists adopting, using, and developing open source software
    - Learning from OSS: particularly, aiming toward sustainability of research computing efforts
    - Clear similarities in the precarious career positions and trajectories of humanware in these two contexts
- Demand more of vendors
- Explore options for collective action/bargaining

Some insights for vendors

- Streamline and otherwise make billing easier for researchers, research support staff, lowly graduate students, etc.
- Be more transparent about:
  - What types of research are and are not being done using [AWS, Azure, GCP]
  - What your development teams are learning from researchers' engagement (and frustrations) with your products and services
    - Recognition and rewards for all types of work, from infrastructure to billing to science
    - Demonstrate how you are working to make responsibilities easier to manage
- Develop strategies for avoiding problems of the past, including:
  - Vendor lock-in/inertia
  - Automating inequalities (h/t Virginia Eubanks)

References

Daniel S. Katz, Jeffrey C. Carver, Sandra Gesing, Karthik Ram, Nicholas Weber (2018). "Planning to Make Research Software More Sustainable via a US Research Software Sustainability Institute (URSSI)" Available at: https://cdn.fs.pathlms.com/1wWrFGH3Te6IrLBD0oEx

Susan Leigh Star & Anselm Strauss (1999). Layers of silence, arenas of voice: The ecology of visible and invisible work. *Computer-Supported Cooperative Work (CSCW)*, *8*(1-2), 9-30.