Computer Supported Cooperative Work (CSCW)
https://doi.org/10.1007/s10606-024-09490-1

RESEARCH

‘The Cloud is Not ***Not*** IT’: Ecological Change in Research Computing in the Cloud

Will Sutherland*[1], Drew Paine[2] & Charlotte P. Lee[1]
*[1] *Human Centered Design and Engineering, University of Washington, Seattle, WA, USA (E-mail: willtsk@uw.edu);* [2] *Lawrence Berkeley National Laboratory, Berkeley, CA, USA*

Accepted: 24 January 2024

Abstract. Along with a number of other computing technologies, cloud computing services are increasingly being promoted as a way of enabling openness, reproducibility, and the acceleration of scientific work. While there have been a variety of studies of the cloud in terms of computing performance, there has been little empirical attention to the changes going on around cloud computing at the level of work and practice. Through a qualitative, ethnographic study, we follow a cosmology research group's transition from a shared high performance computing cluster to a cloud computing service, and examine the cloud service as a coordinative artifact being integrated into a larger ecology of existing practices and artifacts. We find that the transition involves both change and continuity in the group's coordinative work and maintenance work, and point out some of the effects this adoption has on the group's larger set of practices. Finally, we discuss practical implications this has for the broader adoption of cloud computing in university-based scientific work.

Keywords: Artifact Ecologies, Coordinative Artifacts, Cloud Computing, Infrastructure, Research Computing

1 Introduction

Increasingly cloud computing services are promoted as a promising method for providing broader data and compute access, support for reproducibility, and reduced IT work for researchers across the sciences (National Science Foundation (NSF) 2018, n.d.; National Institutes of Health (NIH) 2019; Smith et al. 2019; 2i2c; Brown 2023). Adopting cloud computing is often seen as a way to “accelerate” science, increasing the speed of advances and new discoveries (National Science Foundation (NSF) 2018; Towns et al. 2014; Fortunato et al. 2018), or to “open” science, (National Academies and of Sciences, Engineering, and Medicine 2018; Gentemann et al. 2021; Fecher and Friesike 2014; Vicente-Saez and Martinez-Fuentes, 2018). As Mosconi et al. (2019) argue, these kinds of grand visions for the future of science are likely to have organizational, cultural, and infrastructural consequences, but it hard to know how they will play

Published online: 14 March 2024

out in everyday practice. Research groups are complex sites of work, and it is difficult to predict how the introduction of new technologies such as cloud computing will alter established practices, existing technologies, as well as educational and workforce training activities.

CSCW infrastructure researchers have long explored technology-in-practice in the context of ecologies of technologies, practices, and object relations (Karasti 2014; Cohn 2016; Jackson 2019). Prior work has, for instance, examined changes in the work of ecologists (Jackson and Barbrow 2013) and seafarers (Kongsvik et al. 2020). Among the grand visions for science, CSCW is well-positioned to contribute more sociotechnical accounts, grounded in research practice. Commonplace measures in research computing for cloud computing services, such as compute hours leveraged or counts of studies that can be replicated computationally, can be helpful for scientific researchers, managers, and policy makers. Yet these perspectives consistently provide insufficient insight on how the actual day to day work — the actual practice of science (and scientific training) — is impacted in both local group and global infrastructural contexts.

In this paper we investigate the adoption and use of cloud computing services in a research group in the field of cosmology. Using longitudinal ethnographic data, collected across 9 years, we follow the breakdown of a local computing cluster within this research group, and their subsequent transition to Amazon Web Services (AWS). AWS is an example of an 'on-demand' model of cloud computing that provides access to virtualized or abstracted computing resources on request from a user and charges the user based on usage. We use the notion of *artifact ecologies* (Bødker et al. 2017) to track and analyze shifts in multiple practices carried out across collections of computational resources which are leveraged together within the group. We investigate how this group alters their essential work practices while adopting this on-demand model of cloud computing and integrating its attendant artifacts into their artifact ecology. Specifically we ask: *How does a cosmological research group change their coordinative practices in transitioning to a model of on-demand cloud computing?*

The artifact ecologies perspective (Bødker and Klokmose 2012) encourages us to be specific about changes in practice and technology while simultaneously investigating the breadth of impacts (expected and unexpected) of the adoption of a new technology. We explain two key aspects of how coordinative work changed over the course of the transition and impacted these cosmologist's ongoing work: 1) the re-bounding of the group's field of work, and 2) the re-constitution of testing and maintenance practices. Connecting our findings with the concept of artifact ecologies also gives us a more sophisticated view of how the revolutionary visions projected ahead of a new tool for scientific work are reconciled with established practice and with an existing ecology of instruments and tools. We find that forms of innovation or alteration carry on alongside continuities with past ways of working. This ambiguity is best captured by one lab member's statement that "the cloud is not *not* IT": the IT work of maintaining a cluster did not vanish, but carried on in new forms in working in the cloud. Our observations also provide practical takeaways for understanding networks of maintenance work, lock-in, and the acceleration of scientific work around cloud computing.

2 Literature Review

The literature focused on evaluating and developing cloud services for scientific use overlaps very little with more general discussions of coordination in CSCW, despite verging on some of the same topics. As some members of the CSCW community may not be familiar with on-demand models of research computing we introduce that literature here at some length to contextualize this research in broader concerns before discussing the concepts of coordination that will be employed for our analysis.

2.1 On-demand Models of Research Computing

The "cloud" considered here is a model of provisioning computing resources through a service, and specifically through the collection of artifacts (e.g. configuration files and machine images) which are used to engage and interact with that service. This model of on-demand cloud computing is increasingly being promoted by many funding agencies and researchers as the way forward for most university-based scientific computing. This includes the development of non-commercial national or university-based cyberinfrastructure projects, such as the Aristotle Cloud Federation (Knepper et al. 2019) or the NSF's Jetstream (Stewart et al. 2015), the broader adoption of commercial cloud providers (e.g. Smith et al. 2019; Gentemann et al. 2021), and efforts to provide institutional support for the adoption of such services (National Institutes of Health (NIH) 2019). Commercial cloud computing companies have been actively promoting usage in the sciences, with AWS freely hosting "high-value cloud-optimized datasets" (Amazon Web Services Inc. n.d.a), and touting its ability to "accelerate the pace of innovation" (Amazon Web Services Inc. n.d.b) and allow researchers to "focus on science, not servers" (Amazon Web Services Inc. n.d.c). Microsoft and Google have similarly advertised their usefulness for science, and publicized high-profile discoveries accomplished with their services (Google LLC, n.d.a; Google LLC, n.d.b.; Microsoft Azure, n.d.).

Proponents' arguments for the benefits of cloud computing are various (Table 1). Some of these arguments are about efficiency: that it provides a more efficient rhythm of investment and obsolescence for computing resources (Table 1, #7), that it allows researchers to focus less on data management and IT and more on their primary scientific goals (Table 1, #4), and that it enables "data proximate" computing, which reduces data transfer and allows for the

Table 1 Benefits and challenges to using cloud computing described in the literature.

#	Potential Benefits Benefit	Rationale	Sources
1	Storage and computing efficiency	Can move code to the data (data proximate computing) rather than making copies of large datasets and moving them around. Data storage and retrieval can be optimized on a centralized repository	Bottum et al. 2017; Abernathey et al. 2021; Gentemann et al. 2021
2	Facilitating reproducibility	Reproducibility needs data, code, and software environment to be controlled. Moving away from local file system dependencies helps with this. Storing and sharing infrastructure as code makes it easier to re-create or reconfigure computing environments	Abernathey et al. 2021; Abernathey et al. 2021; Smith et al. 2019
3	Broader access	Having large scale computing resources at one's institution is not a requirement for accessing data because data is available and processable in a central cloud storage location	Bottum et al. 2017; Abernathey et al. 2021; Gentemann et al. 2021; Abernathey et al. 2021
4	Reduces "time to science"	Cloud services handle the maintenance of computing hardware, as well as some operating system configuration work. This reduces the amount of IT or maintenance work researchers must do, as well as hiding "platform complexities" (Bottum et al. 2017), allowing researchers to focus on their primary research concerns	Gentemann et al. 2021; Smith et al. 2019; Bottum et al. 2017
5	Flexibility of infrastructure	Researchers can draw on computing hardware that is maintained and updated continually through a cloud service, rather than investing in hardware at multi-year intervals which then ages and becomes obsolete	Smith et al. 2019; Armbrust et al. 2009
6	Generalization of data management techniques	Working on generic computing platforms will enable researchers to share solutions to common data or computing management issues	Bottum et al. 2017; Smith et al. 2019
7	Elasticity	A cloud service can provide a large amount of computational resources when needed, but is not provisioned and idle when it is not needed	Bottum et al. 2017; Armbrust et al. 2009; Smith et al. 2019

Table 1 (continued)

	Potential Challenges		
#	Barrier	Rationale	Sources
8	Service complexity	Using a cloud computing service requires its own expertise in configuring virtual machines, security, storage, and other technical concerns. Training (and re-training) people in these systems is a significant undertaking	Bottum et al. 2017; Gentemann et al. 2021; Abernathey et al. 2021
9	Friction with current funding models	The rhythms of multi-year grant funding and university overhead costs favor investment in hardware cluster rather than computing as an operational cost. It can be difficult to estimate cloud computing costs over time	Bottum et al. 2017; Abernathey et al. 2021
10	Overheads of cost monitoring	Monitoring expenditures on a cloud service requires work in the form of examining bills and reading up on cloud service policies	Bottum et al. 2017; Gentemann et al. 2021; Sholler 2019
11	Vendor lock in	Because of technology choices or infrastructure development on a particular platform, researchers might find themselves committed to a single vendor. Switching to a vendor for pricing reasons or to take advantage of a better service would then require a great deal of work and potentially cost money in data migration	Armbrust et al. 2009; Smith et al. 2019
12	Data storage and transfer costs	The cost of storing data or moving it around (in particular "egress" costs) on the cloud may be prohibitive	Bottum et al. 2017; Armbrust et al. 2009; Smith et al. 2019
13	Integration or licensing friction	It may be difficult to make use of licensed technologies or identity management systems on a third party system	Bottum et al. 2017; Smith et al. 2019

optimization of data storage in one centralized location (Table 1, #1). Other benefits have to do with coordination or organization of research work, such as facilitating reproducibility (Table 1, #2), and the democratization of access to data (Table 1, #3). These calls for the adoption of the cloud also point to a couple of potential drawbacks or obstacles, particularly the difficulty of learning to use cloud services (Table 1, #8), the potential for vendor lock-in (Table 1, #11), and a mismatch or lack of familiarity between the on-demand computing model and contemporary funding models (Table 1, #9).

Empirical research on the cloud has tended towards testing or benchmarking the architectures and cost structures of cloud services. This involves testing particular kinds of research workflows on cloud architectures or quantifying the actual costs of these workflows on the cloud versus others kinds of HPC resources (Almes et al. 2015; Berriman et al. 2013; Kondo et al. 2009; Deelman et al. 2008; Ramakrishnan et al. 2011; Yelick et al. 2011). A few studies have begun to examine the organizational and work-level dynamics of cloud computing. Boscoe's (2019) ethnographic study of a black hole cosmology group noted how these researchers expressed interest in using the cloud but worried about data transfer costs and hesitated to hand over their closely held data to a commercial entity. Sholler (2019) characterizes some of the necessary but invisible work that comes along with adopting the cloud, such as parsing billing documents. Byrne and Jacobs (2021) highlight unexpected costs, and the difficulty of predicting costs. Other studies using the term "humanware" have emphasized the importance of human expertise in facilitating the adoption and use of cloud services (Song et al. 2019; Voss 2019). Work examining university-level and national-level computing services similarly raise concerns with providing trained technical expertise to assist in onboarding scientists to the cloud (Knepper et al. 2019; Stewart et al. 2015; Toor et al. 2017). Smith et al. (2019) also points out the need for education on cloud technologies and institutional buy-in to motivate a broader move to the cloud amongst astronomers.

The outlook on cloud computing presents a lot of potential outcomes, but with a particular empirical gap. Cloud computing is understood as promoting the acceleration and opening of science, but explicit rationales for this are often not provided, and where explicit connections are made we need better empirical views on how exactly they play out. Cloud computing is understood as promoting or facilitating open science by serving as a platform for providing broader, more democratic access to computing resources (Table 1, #3), simplifying the sharing of data management techniques (Table 1, #6), and facilitating reproducibility (Table 1, #2) (National Academies of Sciences, Engineering, and Medicine 2018; Gentemann et al. 2021; Abernathey et al. 2021). Cloud computing is understood as accelerating science because it allows for the optimization of data storage (Table 1, #1), the reduction of overhead or IT work (#4), and because of its ability to scale (Table 1, #7) (Towns et al. 2014; Gentemann et al. 2021; Abernathey et al. 2021). We cannot ascribe the acceleration of science to faster computing performance or to the reduction of IT work in a straightforward way (for reasons discussed further in the discussion), and the problem of lock in (Table 1, #11) could become an obstacle to promoting openness (Gentemann et al. 2021). But we also cannot simply discount the visions for research cloud computing as boosterism. Many calls for the shift come from researchers themselves, and from practitioners with a great depth of experience in sites of research work. Moreover, prior work has demonstrated how taking up new research tools can indeed have wide-reaching implications for the ordering and organizing of research work (Hine 2006; Thomer and Wickett 2020). Empirical attention is therefore needed, both to better understand the difficulties and opportunities that cloud computing services actually present, as well as the ramifications they might have for research practice.

2.2 Sociotechnical Change and Artifact Ecologies

There are variety of concepts in CSCW and related fields which scholars have used to understand complex sociotechnical change. Studies of infrastructure (Star and Ruhleder 1994) have looked at processes of development (Hanseth and Lyytinen 2008), flexibility over time (Ribes and Polk 2014), and decay (Cohn 2016), as well as longitudinal dynamics such as path dependency (Edwards et al. 2007). Infrastructure studies’ attention to the long term has in fact generated a variety of such dynamics. Shifts across varied infrastructural elements produces tensions, such as that between technological change and ready-to-hand stability, and between development for local uses versus “hardness” towards other use cases (Karasti et al. 2006; Ribes and Finholt 2009). Similar tensions arise between the commitment to standardization and to flexibility (Hanseth et al. 1996). Actors in these large-scale developments are also often managing development (and resources) across multiple evolving projects (Bietz et al. 2013).

The gerund infrastructuring (Neumann and Star 1996; Karasti 2014) aims our attention at infrastructure as an ongoing, processual accomplishment: the alignment of diverse stakeholders, the intentional management of an installed base, and considering responsibility over the long term are aspects of infrastructural change that people carry out on a day-to-day basis (Karasti and Syrjänen 2004). Other studies have similarly located long-term sociotechnical change in ongoing, day-to-day metawork (Neang et al. 2021) and repair (Jackson 2014, 2019). In repair in particular, the work of restoring working order in the wake of breakdowns is generative of new sets of relations between people and their essential resources (Henke 1999; Henke and Sims 2020). The “restored” situation is both a continuity, a return to working order, and also an alteration.

The concept of artifact ecologies can similarly help us understand the breadth and complexity of change that accompanies the adoption of new tools. The concept of artifact ecologies was introduced by Jung et al. (2008) and has subsequently been developed on an activity theoretical basis (Bødker and Klokmose 2011; Korsgaard et al. 2022). It has alignments and divergences with a variety of other conceptions, such as assemblages, surveyed by Lyle et al. (2020). The idea of an artifact ecology renders a given artifact in terms of its shifting relations with other artifacts that a person or a community leverages to accomplish their goals. It brings analytical attention to artifacts as groups, and along with it some mapping methods for untangling and understanding clusters of artifacts and their interactions (Jung et al. 2008; Bødker et al. 2017). Bødker and Klokmose (2012) describe the "stirring up" of an artifact ecology as a new artifact is taken up in place of another, and subsequent equilibrium as possible uses of a device are sussed out by the individual and new ones are improvised. This reworking of activities around the new artifact results in the person using some other artifacts less frequently or abandoning them entirely. *One of the essential elements of the ecology model that we adopt here is that taking up an artifact is not a 1-to-1 replacement of one resource for another, but rather has a ripple effect through a collection of practices and leveraged artifacts.* The design *of* use is a central part of this process (Bødker and Klokmose 2011). People reconstitute practices carried out across a number of artifacts, and change the way they leverage other artifacts in the ecology, finding new uses and changing existing ones.

While our interest in how practices are carried out across clusters of artifacts puts us most closely in line with the artifact ecologies strand of ecologies research outlines by Lyle et al. (2020), we do not draw strongly on the activity-theoretical foundation that underlies much of the research in that area. We are particularly focused on the collection of artifacts the research group leverages in common, in the process of distributed work, rather than an individual's personal ecology of artifacts. For this reason we use the concept in conjunction with the notion of a field of work (Schmidt and Simone 1996), the collection of artifacts through which interdependent work is coordinated.

3 Research Site and Methods

3.1 Research Site

Our findings are based on ethnographic field work with a research group (hereon "the Radio Group") at the University of Washington in the United States. The group's work is in the field of cosmology, and centers around attempts to detect extremely faint emissions from the early universe. This work centers around a number of large, long-running telescope projects, and requires collaboration with other research groups involved in those projects. The Radio Group itself consists of one PI, a research scientist, and 3–4 Ph.D. students. However, members of the group routinely work with members of other labs involved in the telescope projects, attending telecons, having software development meetings, and communicating with other members of their larger collaboration over a messaging platform. The group’s local workspace is a small laboratory setting, which serves as a venue for weekly lab meetings, and houses desks for Ph.D. students. While the room contains a large amount of signal processing hardware, the group carries out their work almost exclusively on their computers and on a large whiteboard, which facilitates discussion of the mathematics and software of the group’s processing pipeline.

Successfully sensing these faint emissions from the early universe would allow researchers to better characterize the “cosmic dawn”, the period during which the first stars and galaxies formed, but it requires extremely fine-tuned signal processing and data analysis techniques. The group’s primary work, therefore, consists of running and plotting data analyses, with the goal of identifying instrumental effects or “systematics” in their instrument, tracking down other software or hardware bugs, or modeling physical phenomena. These tests were usually carried out by Ph.D. students and research scientists and then discussed at lab meetings or over their messaging platform. Discussions around the tests and the plots they produced would then prompt new tests to be carried out (discussed in Paine and Lee 2017).

The group has used a variety of different computing resources for data processing tasks that are too large for local machines. Often these were computing clusters associated with particular institutions or funded through particular research grants. Their primary source of computing for most of the second author’s period of observation was a shared computing cluster located in another laboratory at a different university, funded through a multi-lab research project. Members of the group also used other computing clusters in the course of their work on different projects, including one located with a collaborating research group at the Arizona State University, one housed with the National Radio Astronomy Observatory (NRAO) in the US, and a similar government-operated cluster in Australia. The clusters therefore served as places to store large amounts of data produced by interferometers, as well as a source of computing power necessary to reduce and process that data to produce plots or other analysis outputs. Some of these clusters served this purpose primarily within a single lab, whereas others served researchers across multiple labs, or, in the cases of clusters at government facilities like the NRAO, they served a wider population of radio astronomers. For the purposes of this paper, we will refer to the AWS system as the “cloud” or “cloud service”, the clusters located with specific research groups (and managed by them) as “managed clusters”, and clusters housed at research institutions such as the NRAO as “institutional clusters.”

The group took up the cloud primarily because the computing cluster that they had previously been using was disassembled by the research group who housed it because it had reached the end of its operational lifetime given in its funding.

AWS and other cloud computing companies provide various services for performing computing jobs on machines which the researchers themselves would not need to maintain and which can be scaled up to extremely large sizes as needed. Other members of the Radio Group's larger collaboration had experimented with using the cloud because of these and other potential benefits (Table 1), but had not adopted it, and it was not in use elsewhere amongst their collaborators. The Radio Group had need of a new computing resource but were drawn to the cloud because of these potential benefits, because they did not then have access to other computing clusters through grants or research projects at the time, and because there were computing "credits" available, provided by AWS, which would allow them to do some computing work before they had to start paying for the service.

3.2 Methods

Our findings are developed out of longitudinal observation (Figure 1) and interviews conducted with the Radio Group between 2013 and 2022. The second and last author negotiated access to the field site in 2012 (after first interviewing the PI in June 2011), and the second author conducted observations from 2013–2017. He also returned to the field site for a month during the time when the Radio Group was losing access to its previous computing cluster and starting to transition to the cloud. The first author conducted observations from 2019–2022. Both authors sat in on meetings within the group, conducted interviews with group members as well as others in their larger collaboration, and kept up with asynchronous messaging communications (e.g., email lists, Slack) between the team. Interviews were conducted in person or over video calls, and transcripts were created through a transcription service.

The longitudinal nature of the engagement allowed the authors to compare field notes and experiences from before and after the transition to the cloud, and it allowed the authors to follow the group through critical interactions where established practice, assumptions, and priorities of those involved were foregrounded and negotiated. As pointed out by a number of scholars, (Edwards 2010; Simonsen et al. 2020; Hahn et al. 2018), infrastructural inversion is a technique carried out by those using infrastructure in the course of maintaining and adapting it. In our case this happened at points of breakdown, such as when their primary computing cluster went offline. It also occurred through discussions over

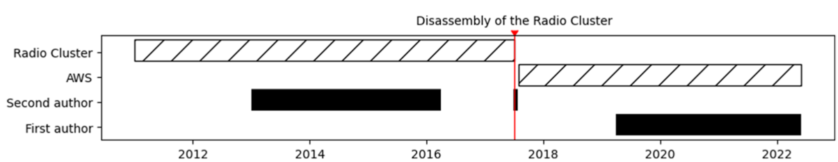

Fig. 1 Timeline showing periods of use of the Radio Cluster and the cloud service along with periods of observation by the authors.

the configuration of schedulers and over strategies for reducing cost on the cloud. These negotiations served as design controversies, through which members of the Radio Group and other labs explicitly surfaced and renegotiated practices for using and maintaining computing resources. These interactions sometimes played out over asynchronous messaging, sometimes through asides in lab meetings, sometimes in in-person meetings, and sometimes in inter-lab video calls. This meant that presence in these different venues was an important part of establishing co-presence (Beaulieu 2010) with the group’s activities.

Cloud computing emerged for the authors as part of the construction of a larger field, focused on coordination around software artifacts as research instruments. The second author examined coordination around managed clusters during their broader studies of research infrastructure in the radio group and in the field of reionization cosmology (Paine and Lee 2017), and the cloud appeared towards the end of that investigation. The first author similarly examined coordination around the cloud service as part of a broader study of research computing. This meant that while our analysis honed in on the cloud service, it was as particular artifacts (e.g. configuration files and “machine images”) that came to be embedded in a larger set of work processes. The analytical and political dimensions of the research object were also shaped by the field (Parmiggiani 2017), as the Radio Group’s transition away from a shared computing cluster shifted our own analytical attention away from a study of cross-lab infrastructure to something closer to a workplace study of the use of the cloud internally within the Radio Group. While our object of study is still most certainly infrastructural, it does suffer from some of the limitations of a single-site implementation study critiqued in Williams and Pollock (2012). The authors made a conscious decision not to extend the field to include activities at AWS due to doubts about access (whether access could be attained but also *where* to gain access if it could be attained) and concerns for scoping labor. Examining AWS’ interactions with research computing stakeholders through an ethnographic position at the company would be an extremely valuable research program, but would deserve more than a little time and effort.

Our analysis is based on interviews (Table 2), fieldnotes collected by the first two authors, messages between group members, documents created by the group, as well as secondary research on AWS and the technologies used on the managed cluster. We conducted open coding of the interviews to identify key concerns of

Table 2 Interviews conducted. Positions are recorded as the individual’s position at the time of the interview. Some individuals are counted for both Ph.D. students and Postdocs because they had transitioned to a new role between interviews.

	Number of interlocutors	**Number of interviews**
Ph.D. Students	6	15
Research Scientists / Postdocs	4	6
Principal Investigators	2	4
	12	25

participants, and used field notes to provide context on events and perspectives. Open coding resulted in 34 codes focused on specific issues in the group's work (e.g. "unexpected costs", "testing cycle", and "cluster politics"). Having identified key concerns and issues, we wanted to know more about how practices changed around those issues. Through discussion, recoding for practices, and lastly comparison with relevant literature, we arrived at the categories presented in the findings: "rebounding the field of work", and "reconstituting maintenance and testing practices". In this redirection in our coding process we also moved from evaluating the challenges or the benefits of the cloud service on its own towards considering change in practice in the research group overall. We drew particularly on studies of ecologies to push our final analysis closer to a holistic consideration of dynamics across all of the group's tools: how technologies and other actors entered and left the ecology, and how practices were reconstituted across these technologies.

This analysis concurrently with and informed further data collection. The first author pursued follow up interviews or asked questions in lab meetings to fill gaps in understanding as well as gather other perspectives on specific issues. The first author also created diagrams of the group's ecology of computing resources as an elicitation technique during two of the interviews with participants who had worked directly on those resources. Interviewees' corrections and responses served as a way of exploring the interrelationships of the artifacts and the practices of maintaining them. Lastly, we followed up with members of the lab by discussing some of the issues described here in lab meetings, and by requesting feedback from them on pre-publication versions of this paper.

4 Findings

We analyze the Radio Group's adoption of the cloud service through two aspects: 1) the re-bounding of a field of work, and 2) the reconstitution of maintenance and testing practices. We center our analysis around the practice of running tests. "Tests" were iterative computing jobs aimed at exploring patterns in the group's data, with the goal of troubleshooting and refining their analysis pipeline. Through relatively small tests, group members would produce plots, which were examined collectively at lab meetings, and based on these discussions the group would plan further tests. Very large scale computing jobs were rare, being carried out usually for a student's dissertation work, and jobs of varying sizes were run on different computing resources: personal laptops, two larger desktops located in the lab, institutional clusters managed by other research groups or research organizations, and their own managed cluster. Early in our observations most larger computing jobs were run on a managed cluster physically located at another university (hereon the Radio Cluster). For the sake of clarity, we highlight a number of artifacts that were leveraged in running these computing jobs on the Radio Cluster, as well as on AWS. We highlight these objects for the sake of explanation, and because they were the primary artifacts through which coordinative work was carried out on the respective systems.

4.1 Key Artifacts of the Radio Group

On the Radio Cluster (Figure 2 and Table 3), computing jobs were run on a shared, live operating system, which was updated and maintained by graduate students and other researchers from different labs who shared access to the cluster. The first considerable object of coordinative work was the software that needed to be installed on the system. This included the group’s analysis software and its dependencies – the other software packages the analysis software relies on to run – as well as other pieces of supporting software, such as databases. These required work in installation and configuration in order to keep them up to date. A second kind of important artifact on the Radio Cluster was the wrapper

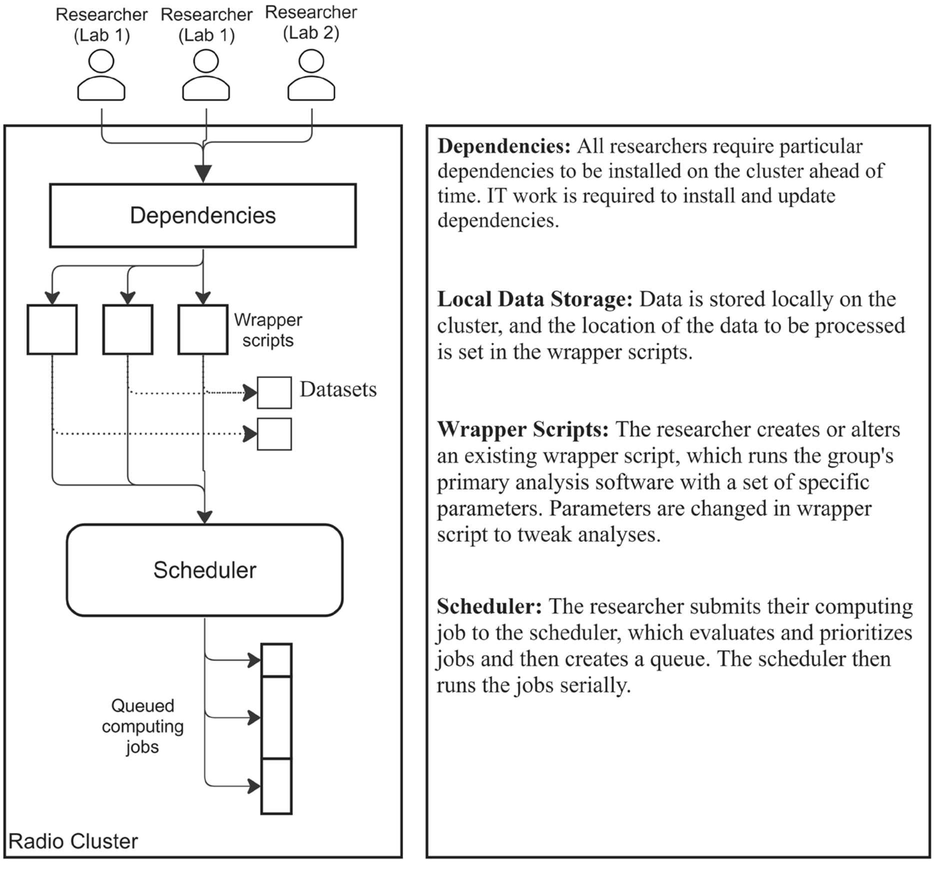

Fig. 2 Simplified, schematic diagram of running a test on the Radio Cluster.

Table 3 Different concerns at issue in the use of artifacts on the Radio Cluster.

Artifact	Negotiated Concerns
Dependencies	Ongoing updates to analysis software and the software on which it depends
Local data storage	Sharing location with other researchers, monitoring storage space
Wrapper Scripts	Specific parameters relevant to an analysis, the location of necessary data products and outputs
Scheduler	Resource requirements for an analysis, fair allocation of time between researchers and placement in queue

script. Wrapper scripts would call the groups' analysis software with particular parameters set, and served as a way of defining and sharing complex parameter combinations for a particular analysis run. The wrapper script could therefore be altered to tweak an analysis in a wide variety of ways, depending on the question that the researcher was attempting to address. Lastly, the Radio Cluster used a scheduler, which would accept job requests from different researchers, organize them into a queue and assign priority based on set criteria, and then schedule the jobs to be run in order. Configuring the criteria for this process and the general behavior of the scheduler was another form of IT work that was occasionally required. Moreover, as we will see, there were sometimes negotiations over the proper configuration of the scheduler.

On AWS (Figure 3 and Table 4), specifications for the size and number of virtual computers could be stored as files, and then used to deploy virtual machines when a member of the lab had a test to run. The most central artifact in this model was the machine image. A machine image is a stored 'snapshot' of a virtualized computer, which can be used to generate new virtualized computers when needed. A computing environment with the group's necessary dependencies could be configured, and then saved as a machine image, allowing members of the group to use this machine image as a kind of template to launch new virtual machines when needed for a computing job. Machine images were configured and pre-loaded with the group's necessary dependencies, and then it could be used to spin up a running virtualized computer (an "instance") on the cloud to do a given computing job. Another central artifact was the cluster configuration file, which specified the number and size of virtualized computers to generate. When used along with the machine image, it could be deployed to generate clusters of varying sizes and costs. This meant that it had to be configured in relation to the concerns of cost and the speed with which the job needed to be done. Wrapper scripts, which were used on the Radio Cluster, were used on AWS in much the same way, being deployed on running computing instances, and used to run a test with a very specific set of parameters. Whereas the machine images were shared, wrapper scripts were more personal to researchers and to specific analyses, but wrapper scripts were in some cases shared between researchers in order to show others particular configurations or the use of particular parameters. Lastly, the

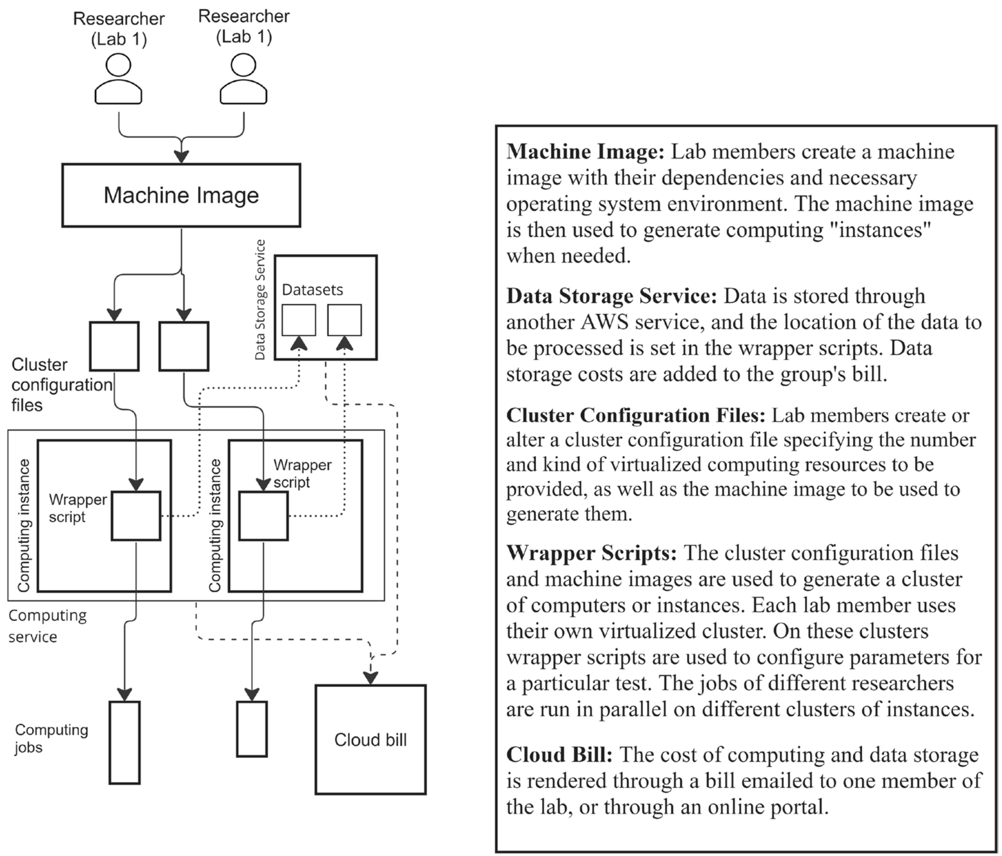

Fig. 3 Simplified, schematic diagram of running a test on AWS. The diagram does not show all artifacts involved, but focuses on those analyzed here.

Table 4 Different concerns at issue in the use of different artifacts on AWS.

Artifact	Negotiated Concerns
Machine images	Continual updates to analysis software and the software on which it depends
Data storage service	Sharing locations between group members, monitoring accumulating storage costs
Cluster configuration files	The cost of computing and the time required for a job
Wrapper scripts	Specific parameters relevant to an analysis, the location of necessary data products and outputs
Cloud bill	Evaluating ongoing costs of data storage and computing jobs, establishing that costs are expected and intentional

cloud service produced a bill, as well as charts that were accessible through an online portal, which indicated how much money had been spent on computing jobs. This itself became an artifact around which the group needed to communicate in order to monitor costs. Taken all together, computing resources on the cloud were leveraged in the manner of templates, which are copied and customized, rather than as a shared environment.

These are simplified accounts of both systems, and there were of course a great many more artifacts and components involved in both systems. We have scoped our analysis here for the sake of clarity, and because these artifacts were the most significant points of negotiation and coordination. With these different artifacts on the two systems established, we turn to the coordinative transition in which they were significant actors.

4.2 Re-bounding the Field of Work

Moving from the Radio Cluster to a cloud service required engaging new people in their ongoing activities, and altering the set of artifacts through which these people interact. The Radio cluster was a managed cluster physically located at another university, and so moving to the cloud extricated the Radio Group's maintenance and testing work from coordinative interactions with other researchers. It also required them to negotiate new working relationships with development processes at AWS, and with people who might provide some technical advice about working on the cloud. They also stopped using certain artifacts associated with the Radio Cluster, and had to take up new ones associated with AWS. It is this process that we describe as the re-bounding of their field of work.

As a shared resource, the Radio Cluster had been the object of significant collaborative maintenance work and negotiation. This occurred in particular around the job scheduler. In an early case, a student at another lab began bypassing the scheduler (and other people's queued jobs) because it was not configured to allocate memory at the rate his analysis required. This prompted a meeting between members of the different labs in which they discussed different designs for the queue, with varying amounts of resources, and weighed different potential shapes for the protocol of using the cluster. Working on a shared cluster therefore required a particular kind of metawork in interweaving different testing tasks from different researchers. This coordinative work was delegated to the system’s job scheduler, and in cases of breakdown the materiality of this scheduler was leveraged in renegotiating the scheduling of different researchers’ computing jobs.

Moving to the cloud obviated this coordinative work of delegating (and negotiating the delegation of) computing time. This stood in contrast not only to the Radio Cluster, where computing time was shared with other researchers, but also with many other institutional and managed clusters that the group had used:

> “You don’t have to fight with anyone. I know Sebastian has been running into a lot of issues with the cluster in [collaborator’s university], where just like your job takes two days to finish the top of the queue or something. AWS is awesome. If money is an issue you can figure out who spent what, but there is no time constraint” (Lillian Ph.D. student).

Lillian contrasts working on the cloud with another institutional cluster that Sebastian, another Ph.D. student in the lab was working with, which had long queue times, and, as Sebastian reported to the authors, an opaque algorithm for deciding priority. Extricating the group from these kinds of queues and negotiations over allocation of computing time was perhaps the primary benefit that Lillian (and others in the Radio Group) ascribed to AWS:

> First Author: “How does that ability make the work different from working through the [Radio cluster]?”
> Lillian (Ph.D. Student): “It’s just faster. The [Radio Cluster] worked really well once everyone had abandoned it, except three people in our group. […] Just everything sped up when you are on the cloud, because you can parallelize it much better.”

Lillian’s term “parallelize” here does not refer to the parallelization of tasks within a computing job, but rather to the fact that different researchers receive distinct computing resources, and are not competing for or negotiating over computing time. The notion of speed here in Lillian’s word “faster” has to do with the group’s ability to quickly get results from tests without having to navigate slow, opaque, and sometimes politically-fraught queueing systems. It is in this sense the understanding of speed was situated in the group’s particular rhythm of testing work.

Coordinative work around the Radio Cluster did not just involve running computing jobs, but also in maintenance work and the risks that came with that. Members of the Radio Group would have to contact people at the university housing the cluster to restart particular machines or perform certain updates. At one point Ivy lost almost all of the data products she had processed on the Radio Cluster due to two A/C outages destroying one of the cluster's disks. Moreover, after the period of use mandated in their grant ended, the group hosting the cluster began to slowly disassemble it, removing machines and using them for other purposes. Finally, they announced that they would completely disassemble the cluster right when Ivy was trying to complete a final analysis for her dissertation. The group was able to negotiate for more time, but had to rapidly find a new resource to store their data and run their analyses. A research scientist in the Radio Group described this kind of "risk" as one of the major drawbacks to working on the Radio Cluster. Liam, the PI of the group, summarized this sentiment towards the "cruft" of managing one's own cluster, saying "I want to *use* a cluster; the last thing I want is to *have* a cluster" (Liam, PI).

While the transition did extricate the group from these kinds of coordinative work, it also meant that the group had a smaller network of collaborators to draw on for technical support. Lillian described her early efforts to get help learning the cloud service early on in her work:

> "I was sold on it. Then I started working to set it up, and realized it was much more complicated. I realized that there wasn't sort of the usage I expected from academic research groups. I expected that I was gonna start doing this, and all these people who've done it before who are gonna show me how it's done, and help me out. I put out some feelers in the [research computing center], and then they connected me with some groups who had done some cloud computing work. Basically I reached out to them and they had said, 'Yeah. We tried and gave up'" (Lillian, Ph.D. student).

Where the Radio Cluster had been the object of collective maintenance effort, the work of setting up and maintaining the cloud was suddenly isolated to within the Radio Group, and on Lillian specifically. Lillian's expectation that she might be able to draw on other researchers for help comes in part from the previously established process of discussing technical issues with other members of the lab or researchers in other labs.

The transition to the cloud did not just shrink the group's field of work, but also grew it: it required that they engage new actors at their university and at AWS. Lillian sought out assistance from other researchers who had experimented with AWS, as well as staff members at a research computing institute at the University of Washington. The latter were able to point her to which AWS services she might need and put her in contact with a developer at AWS. She was then able to consult the developer about the stability of AWS’ cluster management tool in particular:

> “... and I can’t remember what my specific question was then, but it was mostly like what’s going on with this tool? What is the future of it? And he talked it up a lot, you know like we’re doing this active development, we’re coming out with all these new features. I was like ‘ooh, don’t put in new features, like, fix the old ones please.’ But then sure enough like a month later the Parallel Cluster upgrade came out, and it just works a lot better and it’s a lot less finicky than Cfncluster was” (Lillian, PhD Student).

Lillian’s concerns here demonstrate the way that the group’s maintenance work had become contingent upon development work at AWS to a certain extent. For a while, Lillian had to rebuild machine images whenever software updates were pushed from AWS. Eventually the group also had to abandon their job scheduler because AWS stopped supporting it. This relationship therefore created certain kinds of work for Lillian and other members of the lab, but Lillian also recognized the benefit of the forced migration, as the scheduler was a long out of date piece of software which had not itself been updated in years.

In transitioning to AWS, the Radio Group extricated themselves from certain coordinative relations, and at the same time developed new ones. Adopting the cloud service did not just mean adding one technology to the group’s ecology, but it also meant dropping certain other tools and ending other collaborative relations. In this way the adoption had a ripple effect on other parts of the ecology.

4.3 Reconstituting Maintenance and Testing Practices

The transition to the cloud did not just involve altering the boundaries of a field of work, but also reworking the coordinative mechanisms that constituted that field of work. The group meshed new artifacts with existing practices, and in the process altered those practices. This occurred through two processes: 1) the metawork of interweaving the tasks of maintenance and cost monitoring with the work of running tests, and 2) Lillian’s local articulation work in making essential artifacts ready-to-hand for other lab members.

The various artifacts of AWS first had to be worked into the rhythms of maintenance and training in the Radio Group. Larger IT projects, such as setting up AWS, were not delegated at random in the group. Rather, it had been established practice in the group for new Ph.D. students to take on more “IT work” early in their time in the group, before moving on to more research-oriented activities associated with their dissertation work. This kind of service work was a process for delegating maintenance work in a way that balanced the necessity of that work with the scarcity of students’ time and labor. It was in precisely this capacity that Lillian became the “point person” for setting up and maintaining the cloud. Ivy, a more senior student who had done significant amounts of set up and configuration work on the Radio Cluster, described how the work of adopting the cloud fell on the newer students:

> "There was this kind of umm, a little bit of a talk we had in terms of I've done my time. I've done my duty. My volunteer work in getting the [Radio Cluster] working. I spent a lot of time with scripting, GridEngine stuff, umm database stuff. I've done a lot of things that I don't think are terribly fun. I'm not gonna do that for AWS. I've done my time […] so Lillian's on it. She's working on it, umm and she has money, and she's about to do kind of a full test of all the scripting stuff that she's been working on…" (Ivy, Ph.D. Student).

Ivy here refers to a discussion she had with Liam, the PI, and a research scientist in which they made clear that she would not have to take on the sudden IT work of moving to the cloud. In the same way that Ivy here describes having "done my time", Lillian took on the work of setting up and maintaining the cluster as a kind of service work within a division of labor in the lab.

In this capacity as "point person" Lillian had to perform local articulation work in order to make (and maintain) the cloud as a ready-to-hand resource for the group's testing work. A central aspect of this was configuring machine images and cluster configuration files that other members of the group could then take and customize for specific tests. The result of this local articulation work was that the group had a machine image pre-loaded with their essential dependencies, and a cluster configuration file and wrapper scripts which they could alter in order to perform a given test. Lillian would occasionally have to rebuild these images to incorporate updates to the group's software and its dependencies. She also helped train other students in how to use leverage these new computing resources. The first author sat in on a training session in which Lillian instructed two new students on how to go through the workflow she had created: where they could change particular settings for their specific run, and what kinds of breakdowns to avoid. Lillian also outlined this process in a text document explaining each step and where settings should be changed. She also regularly answered questions posed in lab meetings or on Slack when other group members had problems. In this way new artifacts associated with AWS were meshed with established ways of delegating IT work, and Lillian built new practices for maintaining images and cluster configuration files as ready-to-hand computing resources.

Importantly, these practices had to be worked out across a set of interacting artifacts. This is most visible in the case of the group's IDL license. The Radio Group's primary software package had been written a number of years previously in IDL, a proprietary programming language. On the Radio Cluster a license was fairly unproblematically installed on the system, but in setting up the cloud the group had to figure out how to automatically populate these licenses on a number of programmatically-generated computing instances by connecting with a university server. They were eventually able to arrange this with some assistance from the University's IT department, but their solution also occasionally broke down during our observations when the IT department change policies around license-provision. New artifacts had to be aligned and re-aligned with older artifacts in the group's ecology this way in order to *collectively* support practices of testing and maintenance.

This was not a process of fitting new artifacts into existing practices that themselves remained unchanged, but rather the reconstitution of altered coordinative mechanisms around new artifacts. For instance, the concern for computing costs subtly changed the way members of the group ran tests. Running a test on the cloud incurred costs directly to the group's computing funds, and it also carried the risk of unexpected costs. Unexpected costs were incurred, for instance, when the group moved data out of a long term storage space prematurely. There were also accidental costs, such as when large instances were accidentally left on, and one occasion on which somebody scraped the group's access key off of Github and used it to launch a large number of instances (they presumed for the purpose of Bitcoin mining). Avoiding these kinds of unexpected or accidental costs could save the group thousands of dollars, and so part of Lillian's process of becoming comfortable with the cost of on-demand computing was learning to avoid these kinds of pitfalls, and to become confident in spending money when it was necessary:

> "Especially like starting out, when you know, you feel a little bit like paralyzed because you don't want to make a mistake and waste money. And I definitely felt that when I started working on the cloud and completely got over that because I made- I wasted enough money that I am desensitized to it" (Lillian, Ph.D. Student).

When Lillian and other members of the lab set out to run tests on the cloud, they now had to factor the concern for cost into their planning. This introduced a kind of overhead work in thinking about the cost of performing tests, but it also made computing costs difficult to predict. Even with the overhead work of estimating and monitoring costs, Lillian said it was still possible to get it wrong, and there was therefore a need for a "high tolerance" for unexpected costs and a general difficulty in budgeting computing runs.

Incurring expenses was something that members of the lab had to get used to in running tests. Other members of the lab would occasionally express hesitance about running a large test or regret when they had spent money on a test and a bug had prevented them from getting the intended results. During a lab meeting, Sebastian expressed concern over whether some intermediate data products he was storing might be incurring too much expense. Liam and Mila, more senior members of the group, both stated that while that was a good instinct they did not want concern for cost to get in the way of Sebastian using the resources he needed to do his work. In this case more senior members of the lab were intentionally trying to set a tone for managing this concern for cost: encouraging care in using computing resources, but also not letting the concern make students too conservative in running tests and using resources. Through these kinds of interactions as well as through strategies for avoiding unexpected costs, the group managed the issue of computing costs, and returned testing practice to a workable, routine state, but one that involved a new dynamic than it had previously.

In a more general way, cost became the primary functional constraint on the amount of computing the group could leverage, and it caused Lillian to strategize testing around it. This was particularly evident on the Spot Market. The Spot Market is a marketplace for computing instances where currently unused computing capacity can be purchased for significantly less than a dedicated instance, with the caveat that the instance may be "reclaimed" if demand for AWS' computing capacity from other users increases. Lillian pursued this option as a way of saving money, but doing so required working around the interruptions of instances being reclaimed. Lillian noted that this was particularly problematic when running short analyses for testing purposes:

> "I used to use the spot market for everything, and these days I have stopped using it when I am doing like testing one observation, because it was slowing me down a lot when… I don't know if it would be like half the time, probably less than half the time, but some fraction of the time that I ran a test it would terminate because of the spot market and I would have to go back and rerun" (Lillian, Ph.D. student).

This lead Lillian to strategize her analyses between the two services based on the amount of computing needed:

> "… And when tests are like four hours long that [the interruptions] just slows you down a whole bunch so I realized it was just worth the extra money to do testing on non-spot market instances. And then if I am running like 60 observations that'll go on the spot market. If a few of them terminate I can restart them" (Lillian, Ph.D. student).

The cheaper cost of the spot market was worth the interruptions when running longer, more costly tests, but the inconvenience of the interruptions led Lillian to strategize her use of the Spot Market and the on-demand instances to save money at scale but use the dedicated instances (the "non-spot market instances") for short tests. Lillian would "throttle" the size and number of the computing instances she used on the spot market to avoid driving up the price (and therefore causing interruptions). In a lab meeting she pointed out that this felt odd because the whole idea of using the cloud was to make use of its scalability. While the cloud may be limitless in theory, in practice the economics reduced it in size to something larger than the private Radio Cluster but still bounded in everyday use by the amount they were willing to spend and the interruptions they were willing to endure. Members of the lab would also run the smallest tests on local desktop computers in the lab, and analyses associated with particular grants or other institutions could be run on other institutional clusters, for instance at the National Radio Astronomy Observatory (NRAO). In this way the group balanced cost and utility by running different kinds of analyses through different computing resources.

To summarize these points, the process of moving to the cloud involved the integration of new artifacts into practices carried out across *sets* of coordinative artifacts. This further involved the reconstitution of those practices around new dynamics, such as the cost of computing and the work of configuring the number and size of computing resources to be used. By the end of this process, the practice of testing itself looked somewhat like it had previously. The group had restored their practice such that they could run tests "as usual", but this was possible because of significant reorganizations of supporting coordinative practices, especially around maintenance or IT work.

5 Discussion

Examining technologies like cloud services as parts of ecologies can help us understand change in material practice as constrained and shaped by existing ways of working, and having diffuse effects across a web of practices and artifacts. In the case of the Radio Group, taking up the cloud service was not a 1-to-1 replacement of one artifact for another, nor was it a simple addition, with the benefits of one artifact added on top of others. The introduction of AWS into the Radio Group's work had ripple effects on other aspects of their artifact ecology: it involved reworking the relationships between people and artifacts that support their work, as well as altering the nature of their day-to-day practices of conducting research. In this process of change there were important continuities with prior ways of working: while the group performed a great deal of work to maintain the managed cluster, they also performed (different) maintenance work to maintain AMIs and other artifacts on the cloud service. While on a managed cluster they evaluated tradeoffs between computing jobs and time or resources available, on the cloud they evaluated similar concerns against cost. There are a number of other observations we can make when looking at these kinds of transitions as ecological change.

First, it can help us be more specific about what is changing when we talk about changes such as "accelerating" (Towns et al. 2014; Fortunato et al. 2018) science. For the Radio Group the ability to scale up the amount of computing resources at will was certainly a benefit, but the notion of speed that came to matter was in regard to a testing cycle carried out across different computing resources and interwoven with discussions at lab meetings and obstacles created by technical breakdowns and maintenance work. Enabling this testing cycle had less to do with completing single large workloads, and more to do with enabling a rhythm of iterative tests which could inform subsequent tests, a pattern that Byrne and Jacobs (2021) describe as "episodic." In this sense acceleration must be evaluated in terms of the rhythms and temporalities (Jackson et al. 2011; Jackson 2017) of a specific kind of testing work. This is an understanding of speed at the level of practice or routine, to which computing performance or scalability was one significant contributor. While there is a great deal of work on computing performance, there is still relatively little focused on this level (e.g. Goble et al. 2013). Moving to the cloud did reduce the Radio Group's "time to science" (Table 1, #4), but it did so primarily by avoiding opaque queues on each successive testing iteration and extricating them from the coordinative problems of shared cluster maintenance. In regards to the overall maintenance work, the cloud certainly made some things easier, but also introduced new kinds of overheads, such as cost monitoring and the maintenance of machine images. The cloud service also had its own technicity which had to be learned. When the first author prompted the group to reflect on the transition during a lab meeting, they avoided evaluating it in terms of net labor saved. Lillian in particular stated that "the cloud is not *not* IT", emphasizing that there were both breaks and continuities with prior forms of maintenance work.

Second, we can see that the changes occurring around the cloud were diffuse, in that the adoption of the cloud had ripple effects on other parts of the ecology. For the Radio Group, bringing in a new computing resource meant establishing new working relationships (temporary or periodic) with staff at local computing institutes and at AWS, both through direct conversation and through the rhythms of updates pushed to the artifacts associated with AWS. This act of engaging new actors was similar to synergizing work (Bietz, et al. 2010), but it is important to recognize that the group's field of work not only grew, but also shrank in other places. While in this case the group did benefit from extricating themselves from collaborative maintenance of an institutional cluster, they are also unable to benefit from wider coordination of maintenance work (collective troubleshooting, sharing expertise and prior experience) on machine images since their colleagues continue to rely on clusters at universities and institutions. The lack of cross-lab coordination around machine images could certainly be seen as a problem of first adoption, or as one of the "gaps" that open up in the transition between resources (Gentemann et al. 2021). Nevertheless, it makes visible the connection between the adoption of new technological actors and shifts in a field of work. This adds some depth to the problem of platform lock-in (Table 1, #11) in that commitments to new platforms can result in restructuring or segmenting the human infrastructure (Lee et al. 2006) of collaborative maintenance work in unexpected ways. We are not arguing that cloud services will systematically shrink research groups' fields of work, but that individual researchers and larger collaborations will need to navigate this problem of networks of maintenance work being fragmented across different cloud computing service providers or across the cloud and managed clusters. This is particularly true if researchers are to see the benefits of generalizing or sharing computing and data management techniques (Table 1, #6). Provider-agnostic tools (e.g. Terraform) may be another approach to addressing this problem, with the caveat that they would require their own investment of learning and maintenance.

Third, the change we observe here is also diffuse in the sense that its changes were subtle rather than revolutionary. The cloud did not remake the work of the Radio Group entirely, but rather the success of the cloud as a tool lay in the group's ability to align the cloud with other artifacts in established practice. Much of Lillian's work in establishing AWS was in integrating it with the peculiarities of other artifacts (such as accessing IDL licenses) and with existing practices in the lab, such as rhythms of analysis, maintenance, and training work. Moreover, analyses were strategized across a number of different computing resources based on concerns for time and money. These included local laptops, desktops in the lab, other clusters accessible through projects or collaborators, and multiple services on the cloud. In other words, cloud services are best evaluated in how they can be fit into the rhythms of work in a given research group and in the tensions that emerge in using it along with other artifacts and practices essential to that groups' work. This is similar to the process of "mastery" described by (Bødker and Klokmose 2012), in which people gradually gain familiarity and confidence in using an artifact in conjunction with others. In terms of adoption, this means that the usefulness of cloud services will be worked out in interactional alignment (Strauss 1988; Blumer 1969), or in other terms in their ecological flexibility (Luff et al. 1992).

With an ecological view of these kinds of adoption processes, we can be specific about the work dynamics that are changing in a given case, and we can account for the breadth of impact of that adoption, following its expected and unexpected effects on other parts of an ecology. We can also make sense of the often ambiguous character of technological change, in which the adoption of a new tool qualitatively alters many different aspects of people's work in different ways, but net benefits and losses are hard to evaluate. This kind of change can certainly be large or significant, but it is accomplished through compromise and alignment with older artifacts and ways of working.

Lastly, considering the use of cloud computing systems at the level of practice, and in terms of ecologies, makes visible a number of dynamics that are important to consider for researchers transitioning to the cloud as well as policy makers trying to understand how to support such transitions:

1. Working on the cloud may remove or obviate many forms of 'IT work' in maintaining a cluster, but cloud services also have their own technical complexity (Table 1, #8) and require their own overhead work in maintaining images or monitoring cost (see also Sholler 2019).
2. The presence of cost as a concern will not only add extra tasks that need to be done, but can also change computing practices qualitatively because researchers must consider cost when running a computing job or when deciding how long to keep certain data products. Part of training new students or researchers on using the cloud will be training them to avoid cost pitfalls, evaluate different ways of saving money, and become comfortable spending money on computing when they need it. The difficulty of predicting cost on the cloud presents a separate problem on top of overhead monitoring work, which requires research groups to have a higher tolerance for unexpected costs and volatility in computing costs.
3. Researchers will not only need to consider the IT work needed to sustain a cloud service, but also where they will get help and support when things break down. Committing to one platform or another may change who they will have to work with to resolve technical breakdowns (Table 1, #6), whether that be university computing staff, other researchers with one group or another, or whether they are left to their own devices to figure things out from documentation. The Radio Group found the cost of AWS-provided support prohibitive.
4. Moving to the cloud may not be a matter of adopting it as a sole computing resource, but rather it might be taken up as a flexible complement to other resources, used to fill in gaps between projects or to avoid the queues of larger systems for particular kinds of computing jobs. The most productive arrangement of computing in a research group may involve strategizing computing jobs across personal laptops, more powerful lab desktops, cloud services, and institutional clusters to which the group has access through one project or another. This would of course imply taking on the overhead work involved in each system.

Following on this last point, researchers should take these considerations as informing where, when, or for what purposes cloud services might be useful, rather than evaluating them against other kinds of resources as an either/or solution. Similarly, the future of research computing may not be either cloud or not, but rather a changed ecology including cloud services amongst many other resources.

6 Conclusion and Future Work

Our goal in this study has been to examine how a research group alters their essential work practices in adopting a cloud-computing service, and to follow some of the ramifications of this adoption for their larger ecology of research tools and practices. Looking at adoption from this perspective provides practical considerations for researchers and policy makers who are making decisions about cloud computing and need to assess the obstacles to their adoption as well as their consequences and tradeoffs. Fundamentally, we characterize a kind of change around cloud computing resources that is diffuse rather than revolutionary: it has wide-reaching effects on the research group's practices and on their field of work, but also has strong continuities with prior ways of working. While forecasts for technological change in the sciences are clear and grand, the reality of those changes are sure to continue to be ambiguous, involving tradeoffs and unexpected benefits and drawbacks. While many visions of these revolutions focus on a confluence of technologies, rather than the adoption of a single one, as we have examined here, it is nevertheless critical to understand how any arrangements of technologies are integrated into ecologies of existing practices rather than extrapolating sharp revolutionary change from the design of an artifact. The presence of more drawn-out, transformational (Schmidt and Marwick, 2020), and ambiguous changes in scientific practice only increases the need for granular, empirical examinations of change in scientific work. While work on infrastructure, ecologies, repair, and others have begun to engage this problem, it is not solved, and more work is needed to sharpen our analytical tools towards the problem of change explicitly.

6.1 Limitations and Future Work

There are a couple of necessary limitations to this study that future work could target. First, there are a number of issues outlined in Table 1 to which we cannot speak given the context of our study. In particular, the notion of the cloud broadening access was not something that became salient in our observations because the Radio Group's particular type of data is highly tailored to their instrument and is not intended to be shared with a large community outside of their collaboration. The issue of broadening access may nevertheless have significant implications for computing in the sciences, especially in fields such as astronomy where differential access to data is a long-running concern, and more empirical work is needed, particularly in larger projects or collaborations where data access to large numbers of groups is a central problem.

Second, as mentioned in the methods section, our ethnographic field shrank somewhat over the course of the study. What could have been an investigation of computing resources across labs in a larger collaboration narrowed to considering the cloud within the Radio Group only. Expanding the research project to include AWS as a second field site would be valuable. Unfortunately, negotiating access and expanding the scope of research would be time and cost prohibitive so we did not include AWS itself in this study. Future work investigating the cloud as research infrastructure would benefit greatly from perspectives taken across collaborations, from distinct cases, or in following the "biography of the artifact" (Williams and Pollock 2012). It may also allow reflection on some of the other dynamics associated with the cloud, for instance, promoting broader accessibility, the optimization of storage and retrieval, and the benefits for reproducibility.

Acknowledgements

We would like to thank members of the Radio Group for involving us in their work, for their patience with our many questions, and for their comments and thoughts on this publication.

Authors' contributions The first and second authors collected data, all authors contributed to the analysis, and the first author wrote the main text of the manuscript.

Funding This article is based on work supported by National Science Foundation grants (#1954620 and #1302272). Dr. Paine's work at Lawrence Berkeley National Laboratory is supported by the U.S. Department of Energy, Office of Science and Office of Advanced Scientific Computing Research (ASCR) under Contract No. DE-AC02-05CH11231. The views in this paper represent the authors and do not represent those of the U.S. National Science Foundation, Department of Energy, or the University of California.

Data Availability (Not applicable)

Declarations

Ethical Approval This study was approved by the University of Washington Institutional Review Board.

Competing interests The authors declare that they have no competing interests.

References

2i2c. 2023. 2i2c. https://2i2c.org/. Accessed 24 February 2023.

Abernathey, Ryan, Tom Augspurger, Anderson Banihirwe, Charles C. Blackmon-Luca, Timothy J. Crone, Chelle L. Gentemann, Joseph J. Hamman, Naomi Henderson, Chiara Lepore, Theo A. MacCaie, Niall H. Robinson, and Richard P. Signell. 2021. Cloud-native repositories for big scientific data. *Computing in Science & Engineering* 23 (2): 26–35.

Almes, Guy, Celeste Anderson, Curt Hillegas, Tim Lance, Rob Lane, Cliff Lynch, Ruth Marinshaw, Greg Monaco, Eduardo Zaborowski, and Ralph Zottola. 2015. *Research computing in the cloud: Functional considerations for research*. USA: EDUCAUSE Center for Analysis and Research (ECAR), 2015.

Amazon Web Services Inc. (n.d.). AWS Research cloud program. https://aws.amazon.com/government-education/research-and-technical-computing/research-cloud-program/. Accessed 6 August 2023.

Amazon Web Services Inc. (n.d.). Research and Technical Computing on AWS. https://aws.amazon.com/government-education/research-and-technical-computing/ . Accessed 6 August 2023.

Amazon Web Services Inc. (n.d.). AWS Open Data Sponsoship Program. https://aws.amazon.com/opendata/public-datasets/. Accessed 11 April 2023.

Armbrust, Michael, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy H. Katz, Andrew Konwinski, Gunho Lee, David A. Patterson, Ariel Rabkin, Ion Stoica, and Matei Zaharia. 2009. *Above the Clouds: A Berkeley View of Cloud Computing*. University of California, Berkeley, USA: Reliable Adaptive Distributed Systems Laboratory, Electrical Engineering and Computer Sciences Department. https://doi.org/10.1145/1721654.1721672

Atkins, Daniel E., Kelvin K. Droegemeier, Stuart I. Feldman, Hector Garcia-Molina, Michael L. Klein, David G. Messerschmitt, Paul Messina, Jeremiah P. Ostriker, and Margaret H. Wright. 2003. *Revolutionizing Science and Engineering through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure*. Washington D.C., USA: National Science Foundation (NSF), 2003.

Bartling, Sönke, and Sascha Friesike. 2014. Open science: One term, five schools of thought. In *Opening science: The evolving guide on how the internet is changing research, collaboration and scholarly publishing, eds.* Sönke Bartling; *Sascha Friesike*. Springer International Publishing. https://doi.org/10.1007/978-3-319-00026-8

Beaulieu, Anne. 2010. From Co-Location to Co-presence: Shifts in the Use of Ethnography for the Study of knowledge. *Social Studies of Science* 40 (3): 453–470.

Bietz, Matthew J., Eric P. S. Baumer, and Charlotte P. Lee. 2010. Synergizing in Cyberinfrastructure Development. *Computer Supported Cooperative Work (CSCW)* 19 (3–4): 245–281.

Bietz, Matthew J., Drew Paine, and Charlotte P. Lee. 2013. The Work of Developing Cyberinfrastructure Middleware Projects. In *CSCW'13: Proceedings of the 2013 Conference on Computer supported Cooperative Work, San Antonio, Texas, USA,* 23–27 February 2013, 1527–1538. New York: ACM Press.

Blumer, Herbert. 1969. *Symbolic interactionism: Perspective and method*. Oakland: University of California Press.

Bødker, Susanne, and Ellen Christiansen. 2012. Poetry in motion: Appropriation of the world of apps. In *ECCE'12: Proceedings of the 30th European Conference on Cognitive Ergonomics, Edinburgh, United Kingdom,* 28–31 August 2012, 78–84. New York: ACM.

Bødker, Susann, and Clemens Nylandsted Klokmose. 2012. Dynamics in artifact ecologies. In *NordiCHI'12: Proceedings of the 7th Nordic Conference on Human-Computer Interaction, Copenhagen, Denmark,* 14–17 October 2012, 448–457. New York: ACM.

Bødker, Susanne, Peter Lyle, and Joanna Saad-Sulonen. 2017. Untangling the mess of technological artifacts: investigating community artifact ecologies. In *C&T'17: Proceedings of the 8th International Conference on Communities and Technologies, Troyes, France,* 26–30 June 2017, 246–255. New York: ACM.

Bødker, Susanne, and Clemens Nylandsted Klokmose. 2011. The human–artifact model: An activity theoretical approach to artifact ecologies. *Human-Computer Interaction* 26 (4): 315–371.

Borne, K.D., S. Jacoby, K. Carney, A. Connolly, T. Eastman, M. J. Raddick, J. A. Tyson, and J. Wallin. 2009. The revolution in astronomy education: Data science for the masses. arXiv preprint arXiv:0909.3895.

Boscoe, Bernadette Marie. 2019. *From Blurry Space to a Sharper Sky: Keeping Twenty-Three Years of Astronomical Data Alive*. Ph.D. dissertation. University of California, Los Angeles, USA: Dept. of Information Studies, School of Education and Information Studies.

Bottum, Jim, Dustin Atkins, Alan Blatecky, Rick McMullen, Todd Tannenbaum, Jan Cheetham, Jim Wilgenbusch, Karan Bhatia, Erik Deumens, Barr von Oehsen, Geoffrey Fox, Marcin Ziolkowsky, Asbed Bedrossian, and Dan Fay. 2017. *The future of cloud for academic research computing*. Washington D.C USA: The National Science Foundation.

Boujut, Jean-François, and Eric Blanco. 2003. Intermediary objects as a means to foster co-operation in engineering design. *Computer Supported Cooperative Work (CSCW)* 12 (2): 205–219.

Brown, Ben. 2023. DOE's Integrated Research Infrastructure (IRI): Vision, strategy, and implementation. Department of energy office of science. https://science.osti.gov/-/media/ascr/ascac/pdf/meetings/202306/Brown_IRI_ASCAC_2023206.pdf. Accessed 2 August 2023.

Bruce, Berriman G., Ewa Deelman, Gideon Juve, Mats Rynge, and Jens S. Vöckler. 2013. The Application of Cloud Computing to Scientific Workflows: A Study of Cost and Performance. *Philosophical Transactions of the Royal Society a: Mathematical, Physical and Engineering Sciences* 371 (1983): 1–14.

Byrne, Ruby, and Daniel Jacobs. 2021. Development of a high throughput cloud-based data pipeline for 21 cm cosmology. *Astronomy and Computing* 34: 100447.

Cohn, M. L. 2016. Convivial decay: Entangled lifetimes in a geriatric infrastructure. In *CSCW '16: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing, San Francisco*, USA, 26 February - 2 March 2016. New York: ACM.

Deelman, Ewa, Gurmeet Singh, Miron Livny, Bruce Berriman, and John Good. 2008. The cost of doing science on the cloud: The montage example. In *SC'08: Proceedings of the 2008 ACM/IEEE Conference on Supercomputing, Austin, Texas, USA,* 15–21 November 2008, 1–12.

Dirks, Lee. 2009. Introduction. In *The fourth paradigm: Data-intensive scientific discovery,* eds. Tony Hey, Stewart Tansley, and Kristen Tolle. Redmond, WA: Microsoft research.

Edwards, Paul N. 2010. *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*. Cambridge, MA: MIT Press.

Edwards, S. J. Jackson, G. C. Bowker, C. P. Knobel. 2007. *Understanding infrastructure: Dynamics, tensions, and design*. NSF Report of a Workshop on the History & Theory of Infrastructure: Lessons for New Scientific Cyberinfrastructures.

Fecher, Benedikt, and Sascha Friesike. 2014. Open science: one term, five schools of thought. In *Opening science: The evolving guide on how the internet is changing research, collaboration and scholarly publishing,* eds. Sonke Bartling, and Sascha Friesike. Springer International Publishing.

Fortunato, Santo, Carl T. Bergstrom, Borner Katy, James A. Evans, Helbing Dirk, Milojević Staša, Alexander M. Petersen, Radicchi Filippo, Sinatra Roberta, Uzzi Brian, Vespignani Alessandro, Waltman Ludo, Wang Dashun, and Albert-Laszlo Barabasi. 2018. Science of science. *Science* 359:6379.

Fujimura, Joan H. 1987. Constructing `Do-Able' Problems in Cancer Research: Articulating Alignment. *Social Studies of Science* 17 (2): 257–293.

Gentemann, Chelle L., Chris Holdgraf, Ryan Abernathey, Daniel Crichton, James Colliander, Edward J. Kearns, Yuvi Panda, and Richard P. Signell. 2021. Science storms the cloud. *AGU Advances* 2 (2): 1–7.

Gerson, Elihu M. 2008. Reach, Bracket, and the Limits of Rationalized Coordination: Some Challenges for CSCW. In *Resources, Co-Evolution and Artifacts: Theory in CSCW*, eds. M. S. Ackerman, C. A. Halverson, T. Erickson, and W. A. Kellogg, 193–220. London: Springer.

Goble, Carole, David De Roure, and Sean Bechhofer. 2013. Accelerating scientists' knowledge turns. In *IC3K'13: Proceedings of the International Joint Conference on Knowledge Discovery, Knowledge Engineering, and Knowledge Management, Algarve, Portugal,* 19–22 September 2013, 3–25. Berlin: Springer.

Google LLC (n.d.a). Google Cloud supports international effort to produce first-ever image of a black hole. Google LLC. https://edu.google.com/intl/ALL_us/why-google/customer-stories/eht-gcp/. Accessed 11 April 2023.

Google LLC (n.d.b). Cloud Life Sciences. Google LLC. https://cloud.google.com/life-sciences. Accessed 11 April 2023.

Hahn, Charlie, Andrew Hoffman, Steven Slota, Sarah Inman, and David Ribes. 2018. Entangled inversions: Actor/analyst symmetry in the ethnography of infrastructure. *Interaction Design and Architecture* 38 (1, Summer 2018): 124–139.

Hanseth, Ole, and Kalle Lyytinen. 2010. Design theory for dynamic complexity in information infrastructures: The case of building internet. *Journal of Information Technology* 25 (1): 104–142.

Hanseth, Ole, Eric Monteiro, and Morten Hatling. 1996. Developing information infrastructure: The tension between standardization and flexibility. *Science, Technology, & Human Values* 21 (4): 407–426.

Hanseth, O., and K. Lyytinen. 2008. Theorizing about the Design of Information Infrastructures: Design kernel theories and principles. *Sprouts: Working Papers on Information Environments, Systems and Organizations* 4(4): 207–241.

Henke, C. R. 1999. The mechanics of workplace order: Toward a sociology of repair. *Berkeley Journal of Sociology* 44 (1): 55–81.

Henke, Christopher, and Benjamin Sims. 2020. *Repairing infrastructures: The maintenance of materiality and power*. Cambridge, MA: MIT Press.

Hepsø, Vidar, Eric Monteiro, and Knut Rolland. 2009. Ecologies of eInfrastructures. *Journal of the Association for Information Systems (JAIS)* 10 (5): 430–446.

Hey, Tony, Stewart Tansley, and Kristin Tolle, eds. 2009. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Redmond, WA: Microsoft research.

Hine, Christine. 2006. Databases as Scientific Instruments and Their Role in the Ordering of Scientific Work. *Social Studies of Science* 36 (2): 269–298.

Iosup, Alexandru, Simon Ostermann, M. Nezi, Yigitbasi, Radu Prodan, Thomas Fahringer, and Dick Epema. 2011. Performance analysis of cloud computing services for many-tasks scientific computing. *IEEE Transactions on Parallel and Distributed Systems* 22 (6): 931–945.

Jackson, Steven J. 2014. Rethinking Repair. In *Media Technologies: Essays on Communication, Materiality, and Society*, eds. T. Gillespie, P. J. Boczkowski, and K. A. Foot, 221–239. Cambridge, MA: The MIT Press.

Jackson, Steven J. 2017. Speed, Time, Infrastructure: Temporalities of Breakdown, Maintenance, and Repair. In *The Sociology of Speed: Digital, Organizational, and Social Temporalities*, eds. J. Wajcman and N. Dodd, 169–185. Oxford, UK: Oxford University Press.

Jackson, Steven J. 2019. Repair as Transition: Time, Materiality, and Hope. In *Repair Work Ethnographies: Revisiting Breakdown, Relocating Materiality,* 337-347, eds. I. Strebel, A. Bovet, and P. Sormani. Singapore: Springer, Singapore.

Jackson, Steven J., and Sarah Barbrow. 2013. Infrastructure and vocation: Field, calling and computation in ecology. In *CHI'13: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Paris, France,* April 27 – May 2 2013, 2873–2882. New York: ACM Press.

Jackson, Steven J., David Ribes, Ayse Buyuktur, Geoffrey C. Bowker. 2011. Collaborative Rhythm: Temporal Dissonance and Alignment in Collaborative Scientific Work. In *CSCW'11: Proceedings of the ACM 2011 Conference on Computer supported Cooperative Work, Hangzhou, China,* 19–23 March 2011, 245–254. New York: ACM Press

Jirotka, Marina, Nigel Gilbert, and Paul Luff. 1992. On the social organisation of organisations. *Computer Supported Cooperative Work (CSCW)* 1 (1): 95–118.

Jung, Heekyoung, Erik Stolterman, Will Ryan, Tonya Thompson, and Marty Siegel. 2008. Toward a framework for ecologies of artifacts: How are digital artifacts interconnected within a personal life? In *NordiCHI'08: Proceedings of the 5th Nordic Conference on Human-Computer Interaction, Lund, Sweden,* 20–22 October 2008, 201–210. New York: ACM Press.

Kaltenbrunner, Wolfgang. 2017. Digital Infrastructure for the Humanities in Europe and the US: Governing Scholarship through Coordinated Tool Development. *Computer Supported Cooperative Work (CSCW)* 26 (3): 275–308.

Karasti, Helena, Karen S. Baker, and Eija Halkola. 2006. Enriching the Notion of Data Curation in eScience: Data Managing and Information Infrastructuring in the Long Term Ecological Research (LTER) Network. *Computer Supported Cooperative Work (CSCW)* 15 (4): 321–358.

Karasti, Helena, and Anna-Liisa Syrjänen. 2004. Artful infrastructuring in two cases of community PD. In *PDC'04: Proceedings of the Eighth Conference on Participatory Design: Artful Integration: Interweaving Media, Materials and Practices, Toronto, Canada,* 27–31 July 2004, 20–30. New York: ACM.

Karasti, Helena. 2014. Infrastructuring in participatory design. In *PDC'14: Proceedings of the 13th Participatory Design Conference, Windhoek, Namibia,* 6-10 October 2014. New York: ACM.

Kee, Kerk, and Larry Browning. 2010. The Dialectical Tensions in the Funding Infrastructure of Cyberinfrastructure. *Computer Supported Cooperative Work (CSCW)* 19 (3–4): 283–308.

Knepper, Richard, Susan Mehringer, Adam Brazier, Brandon Barker, and Resa Reynolds. 2019. Red cloud and aristotle: Campus clouds and federations. In *HARC'19: Proceedings of the humans in the loop: Enabling and facilitating research on cloud computing, Chicago, USA,* 1-6. New York: ACM.

Knorr-Cetina, Karin. 1999. *Epistemic Cultures: How the Sciences Make Knowledge*. Cambridge, MA: Harvard University Press.

Kondo, Derrick, Bahman Javadi, Paul Malecot, Franck Cappello, and David P. Anderson. 2009. Cost-Benefit Analysis of Cloud Computing Versus Desktop Grids. In *IPDPS'09: IEEE International Symposium on Parallel & Distributed Processing, Rome, Italy,* 23–29 May 2009, 1–12. New York: IEEE.

Korsgaard, Henrik, Peter Lyle, Joanna Saad-Sulonen, Clemens Nylandsted Klokmose, Midas Nouwens, and Susanne Bødker. 2022. Collectives and their artifact ecologies. *Proceedings of the ACM on Human-Computer Interaction*, November 2022, 6(CSCW2):1–26.

Larsen-Ledet, Ida, Henrik Korsgaard, and Susanne Bødker. 2020. Collaborative writing across multiple artifact ecologies. In *CHI'20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems Honolulu, HI, USA*, 25–30 April 2020, 1–14. New York: ACM.

Lee, Charlotte P., and Kjeld Schmidt. 2018. A bridge too far? Critical remarks on the concept of 'infrastructure' in CSCW and IS. In V. Wulf, V. Pipek, D. Randall, M. Rohde, Kjeld Schmidt, Gunnar Stevens (eds): *Socio-Informatics: A Practice-based Perspective on the Design and Use of IT Artifacts*. 177–218. Oxford: Oxford University Press.

Lee, Charlotte P., Paul Dourish, and Gloria Mark. 2006. The Human Infrastructure of Cyberinfrastructure. In *CSCW'06: Proceedings of the 2006 20th anniversary Conference on Computer supported Cooperative Work, Banff, Alberta, Canada,* 4–8 November 2006, 483–492. New York: ACM.

Luff, Paul, Christian Heath, and David Greatbatch. 1992. Tasks-in-interaction: paper and screen based documentation in collaborative activity. In *CSCW'92: Proceedings of the 1992 ACM Conference on Computer-Supported Cooperative Work, Toronto, Canada,* 1–4 November 1992, 163–170. New York: ACM.

Lyle, Peter, Henrik Korsgaard, and Susanne Bødker. 2020. What's in an ecology? A review of artifact, communicative, device and information ecologies. In *NordiCHI'20: Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society, Talinn, Estonia,* 25–29 October 2020, 1–14. New York: ACM.

Microsoft Azure (n.d.). Azure high-performance computing for health and life sciences. Microsoft Azure. https://www.microsoft.com/en-us/research/academic-program/microsoft-azure-for-research/. Accessed 12 April 2023.

Monteiro, Eric, Neil Pollock, Ole Hanseth, and Robin Williams. 2013. From artefacts to infrastructures. *Computer Supported Cooperative Work (CSCW)* 22 (4): 575–607.

Mosconi, Gaia, Qinyu Li, Dave Randall, Helena Karasti, Peter Tolmie, Jana Barutzky, Matthias Korn, and Volkmar Pipek. 2019. Three gaps in opening science. *Computer Supported Cooperative Work (CSCW)* 28 (3): 749–789.

National Academies of Sciences, Engineering, and Medicine. 2018. Open science by design: Realizing a vision for 21st century research. https://doi.org/10.17226/25116

National Institutes of Health (NIH). 2019. About the Strides Initiative. National Institutes of Health (NIH). https://datascience.nih.gov/strides. Accessed 12 April 2023.

National Science Foundation (NSF) 2018. NSF and Internet2 to Explore Cloud Computing to Accelerate Science Frontiers. National Science Foundation (NSF). https://nsf.gov/news/news_summ.jsp?cntn_id=297193. Accessed 12 April 2023.

Neang, Andrew B., Will Sutherland, Michael Beach, and Charlotte Lee. 2021. Data integration as coordination: The articulation of data work in an ocean science collaboration. *Proceedings of the ACM Human-Computer Interaction*, December 2020, 4 (CSCW3): 1–25.

Paine, Drew, and Charlotte P. Lee. 2017. "Who Has Plots?": Contextualizing Scientific Software, Practice, and Visualizations. *Proceedings of the ACM on Human-Computer Interaction*, November 2017, 1 (CSCW): 1–21.

Parmiggiani, Elena. 2017. This is not a fish: On the scale and politics of infrastructure design studies. *Computer Supported Cooperative Work (CSCW)* 26 (1–2): 205–243.

Ramakrishnan, Lavany, Piotr T. Zbiegel, Scott Campbell, Rick Bradshaw, Richard Shane Canon, Susan Coghlan, Iwona Sakrejda, Narayan Desai, Tina Declerck, Anping Liu. 2011. Magellan: Experiences from a Science Cloud. In *ScienceCloud'11: Proceedings of the 2nd International Workshop on Scientific Cloud Computing, San Jose, California, USA,* 8 June 2011, 49–58. New York: ACM.

Ribes, David, and Thomas A. Finholt. 2009. The Long Now of Technology Infrastructure: Articulating Tensions in Development. *Journal of the Association for Information Systems* 10 (5): 375–398.

Ribes, D., and J. B. Polk. 2014. Flexibility relative to what? Change to research infrastructure. *Journal of the Association for Information Systems* 15 (5): 287–305.

Ribes, David, Steven Jackson, Stuart Geiger, Matthew Burton, and Thomas A. Finholt. 2013. Artifacts That Organize: Delegation in the Distributed Organization. *Information and Organization* 23 (1): 1–14.

Ribes, David, and Thomas A. Finholt. 2008. Representing community: knowing users in the face of changing constituencies. In *CSCW'08: Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work, San Diego, California, USA,* 8–12 November 2008, 107–116. New York: ACM.

Ruben Vicente-Saez, Clara Martinez-Fuentes. 2018. Open Science now: A systematic literature review for an integrated definition. *Journal of Business Research* 88: 428–436.

Rule, Adam, Ian Drosos, Aurèlien Tabard, and James D. Hollan. 2018. Aiding collaborative reuse of computational notebooks with annotated cell folding. *In Proceedings of the ACM on Human-Computer Interaction* 2 (CSCW): 1–12.

Schmidt, Kjeld, and Ina Wagner. 2002. Coordinative artifacts in architectural practice. In *COOP'2002: Cooperative systems design - a challenge of the mobility age, Saint-Raphaël, France,* June 4 – 7 2002, 257–274. Amsterdam: IOS Press.

Schmidt, Kjeld, and Carla Simone. 1996. Coordination Mechanisms: Towards a Conceptual Foundation of CSCW Systems Design. *Computer Supported Cooperative Work (CSCW)* 5 (2–3): 155–200.

Schmidt, Kjeld, and Ina Wagner. 2004. Ordering systems: Coordinative practices and artifacts in architectural design and planning. *Computer Supported Cooperative Work (CSCW)* 13 (5–6): 349–408.

Sholler, Dan. 2019. ''Invisible Work'' as a Lens for Understanding Humanware's Role in Research Cloud Computing: Evidence from an Interview-Based Study. In *HARC'19: Proceedings of the Humans in the Loop: Enabling and Facilitating Research on Cloud Computing, Chicago, IL, USA,* July 29 2019, 1–5. New York: ACM.

Simonsen, Jasper, Helena Karasti, and Morten Hertzum. 2020. Infrastructuring and participatory design: Exploring infrastructural inversion as analytic, empirical and generative. *Computer Supported Cooperative Work (CSCW)* 29 (1–2): 115–151.

Smith, Arfon M., Rob Pike, William O'Mullane, Frossie Economou, Adam Bolton, Ivelina Momcheva, Amanda E. Bauer, Bruce Becker, Eric Bellm, Andrew Connolly, Steven M. Crawford, Nimish Hathi, Peter Melchior, Joshua Peek, Arif Solmaz, Ross Thomson, Erik Tollerud, and David W. Liska. 2019. Arxiv. https://ui.adsabs.harvard.edu/abs/2019arXiv190706320S. Accessed 12 April 2023.

Song, Yongwook, Xu Fu, and Chris Richards. 2019. A Use Case of Humanware and Cloud-Based Cyberinfrastructure: Time-Series Data Classification Using Machine Learning. In *HARC'19: Proceedings of the Humans in the Loop: Enabling and Facilitating Research on Cloud Computing, Chicago, USA,* 29 July 2019, 1–7. New York: ACM.

Sonnenwald, Diane H. 1995. Contested collaboration: A descriptive model of intergroup communication in information system design. *Information Processing & Management* 31 (6): 859–877.

Star, S. L., and K. Ruhleder. 1994. Steps towards an ecology of infrastructure: complex problems in design and access for large-scale collaborative systems. In *CSCW'94: Proceedings of the 1994 ACM conference on Computer Supported Cooperative Work, Chapel Hill, USA,* 22–26 October 1994, 253 – 264. New York: ACM.

Star, Susan L., and Geoffrey C. Bowker. 2006. How to infrastructure. In *Handbook of new media: Social shaping and social consequences of ICTs*, eds. L. Lievrouw and S. Livingstone, 230–245. London: Sage.

Star, Susan L., and James R. Griesemer. 1989. Institutional ecology, ‘translations’ and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907–39. *Social Studies of Science* 19 (3): 387–420.

Steinhardt, Stephanie B., and Steven Jackson. 2014. Reconciling Rhythms: Plans and Temporal Alignment in Collaborative Scientific Work. In *CSCW'14: Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing, Baltimore, Maryland, USA,* 15–19 February 2014, 134–145. New York: ACM.

Stewart, Craig A., Timothy M. Cockerill, Ian Foster, David Hancock, Nirav Merchant, Edwin Skidmore, Daniel Stanzione, James Taylor, Steven Tuecke, George Turner, Matthew Vaughn, and Niall I. Gaffney. 2015. Jetstream: A Self-Provisioned, Scalable Science and Engineering Cloud Environment. In *XSEDE'15: Proceedings of the 2015 XSEDE Conference: Scientific Advancements Enabled by Enhanced Cyberinfrastructure, St. Louis, Missouri,* 26–30 July 2015, 1 – 8. New York: ACM.

Strauss, Anselm. 1988. The Articulation of Project Work: An Organizational Process. *The Sociological Quarterly* 29 (2): 163–178.

Thomer, Andrea K., and Karen M. Wickett. 2020. Relational data paradigms: What do we learn by taking the materiality of databases seriously? *Big Data & Society*, January - June 2020, 7 (1): 1–16.

Toor, Salman, Mathias Lindberg, Ingemar Fällman, Andreas Vallin, Olof Mohill, Pontus Freyhult, Linus Nilsson, Martin Agback, Lars Viklund, Henric Zazzi, Ola Spjuth, Marco Capuccini, Joakim Möller, Donal Murtagh, and Andreas Hellander. 2017. SNIC Science Cloud (SSC): A National-Scale Cloud Infrastructure for Swedish Academia. In *eScience'17: Proceedings of the 2017 IEEE 13th International Conference on e-Science (e-Science), Auckland, New Zealand,* 24–27 October 2017, 219–227.

Towns, John, Tim Cockerill, Maytal Dahan, Ian Foster, Kelly Gaither, Andrew Grimshaw, Victor Hazlewood, Scott Lathrop, Dave Lifka, Gregory D. Peterson, Ralph Roskies, J. Ray Scott, and Nancy Wilkins-Diehr. 2014. XSEDE: Accelerating scientific discovery. *Computing in Science & Engineering* 16 (5): 62–74.

Trigg, Randall H., and Susanne Bødker. 1994. From implementation to design: Tailoring and the emergence of systematization in CSCW. In *CSCW'94: Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work, Chapel Hill, NC, USA,* 22–26 October 1994, 45–54. New York: ACM.

Trond Kongsvik, Torgeir Haavik, Rolf Bye, and Petter Almklov. 2020. Re-boxing seamanship: From individual to systemic capabilities. *Safety Science*, October 2020, 130 (1): 1–10. https://doi.org/10.1016/j.ssci.2020.104871

Voss, Brian D. 2019. Humanware: The Critical Role of People in Supporting Research in the Cloud. In *HARC'19: Proceedings of the Humans in the Loop: Enabling and Facilitating Research on Cloud Computing, Chicago, USA,* 29 July 2019, 1–6. New York: ACM.

Williams, Robin, and Neil Pollock. 2012. Research commentary—moving beyond the single site implementation study: How (and why) we should study the biography of packaged enterprise solutions. *Information Systems Research*, June 2012, 23 (1): 1–22. https://doi.org/10.1287/isre.1110.0352

Yelick, Katherine, Susan Coghlan, Brent Draney, Richard Shane Canon, Lavanya Ramakrishnan, Adam Scoval, Iwona Sakrejda, Anping Liu, Scott Campbell, Piotr T. Zbiegiel, Tina Declerck, and Paul Rich. 2011. *The Magellan Report on Cloud Computing for Science*. Office of Advanced Scientific Computing Research (ASCR).