Designing for Reproducibility: A Qualitative Study of Challenges and Opportunities in High Energy Physics

Sebastian S. Feger
LMU Munich and CERN
Munich, Germany and Geneva, Switzerland
sebastian.s.feger@cern.ch

Albrecht Schmidt
LMU Munich
Munich, Germany
albrecht.schmidt@ifi.lmu.de

Sünje Dallmeier-Tiessen
CERN
Geneva, Switzerland
sunje.dallmeier-tiessen@cern.ch

Paweł W. Woźniak
Utrecht University
Utrecht, the Netherlands
p.w.wozniak@uu.nl

ABSTRACT

Reproducibility should be a cornerstone of scientific research and is a growing concern among the scientific community and the public. Understanding how to design services and tools that support documentation, preservation and sharing is required to maximize the positive impact of scientific research. We conducted a study of user attitudes towards systems that support data preservation in High Energy Physics, one of science's most data-intensive branches. We report on our interview study with 12 experimental physicists, studying requirements and opportunities in designing for research preservation and reproducibility. Our findings suggest that we need to design for motivation and benefits in order to stimulate contributions and to address the observed scalability challenge. Therefore, researchers' attitudes towards communication, uncertainty, collaboration and automation need to be reflected in design. Based on our findings, we present a systematic view of user needs and constraints that define the design space of systems supporting reproducible practices.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; **Collaborative and social computing systems and tools**.

KEYWORDS

Reproducible Research, Design Requirements, Interview Study, Secondary Usage Forms

1 INTRODUCTION

Reproducibility and reusability are core scientific concepts, enabling knowledge transfer and independent research verification. Alarming reports concerning the failure to reproduce empirical studies in a variety of scientific fields [2, 12, 45] are leading to the development of services, tools and strategies that aim to support key reproducible research practices [60].

Preserving and *sharing* research are basic requirements in reproducible science [4, 29, 58], requiring efforts to describe, clean and document resources [13]. But those efforts are often not matched by the perceived gain. In fact, studies claim that the scientific culture does not support or even impairs compliance with reproducible practices [5, 21].

As research preservation tools are emerging, we set out to study design requirements for technology that supports reproducible research practices. We studied data sharing and preservation flows and attitudes towards preservation systems in High Energy Physics (HEP), one of the most data-intensive branches of science [30]. The volume of data and the community's demonstrated early adoption of computer-supported technology — most notably the invention of the World Wide Web [9] — make for a strong environment to study technologies and strategies that are expected to become increasingly relevant in data-driven science; also referred to as the fourth paradigm of science [7].

We conducted our interview study with experimental physicists at CERN, a key HEP laboratory. The study was closely connected to a research preservation prototype service, tailored to CERN's major experiments. Based on our

findings, we map practices around data sharing and chart challenges and opportunities involved in designing for research preservation and reproducibility. This paper presents: (1) a detailed description of data preservation flows in world's leading data-intensive science environment; (2) six themes that describe user attitudes towards data presentation systems and (3) implications for designing systems that support reproducible science.

This paper is organized as follows. First, we review requirements and challenges of reproducible research and past efforts in designing for research communities. Next, we describe our study's context, in particular HEP and the prototype research preservation service. We then provide details of our interview study. Afterwards, we report on the six themes we identified: Motivation, Communication, Uncertainty, Collaboration, Automation and Scalability. Finally, we present implications for designing technology that supports reproducible research practices.

2 RELATED WORK

In this section, we provide: (1) an overview of definitions, requirements, discussed incentive structures for reproducible research and reflect on discussions concerning the role of replication in HCI; and (2) review previous work in designing for scientific communities and research practices.

Reproducibility

Definitions of reproducibility and related terms vary between different disciplines [28]. Leek and Peng [39] define reproducibility *"as the ability to recompute data analytic results given an observed dataset and knowledge of the data analysis pipeline."* Feitelson [28] stresses that reproducibility is not limited to simply recreating exactly the same experiment, but defines it as a *"reproduction of the gist of an experiment: implementing the same general idea, in a similar setting, with newly created appropriate experimental apparatus."*

The latter definition of reproducibility fits well to data analysis in HEP, characterized by statistically combining earlier experiment data with later run data. This data enrichment allows researchers to prove scientific concepts based on statistical probability. Since analyses might be based on experiment data captured over a range of several years, the former definition of reproducibility applies: analyses are not simply re-executed, but enriched and adapted to new input.

In this paper, we use the terms reproducibility and reproducible science. While it is important for us to refer to semantic discussions [24, 28, 33] regarding reproducibility and related terms, like replicability and repeatability, we aim generally at environments in which researchers are encouraged to describe, preserve and share their work, in order to make resources *re-usable* in the future.

Description and Preservation are Requirements. In order to enable the reproducibility of an experiment, researchers have to follow a set of practices [4, 13]. Those include documentation of all relevant analysis artefacts. In their paper, Bánáti et al. [3] classified several dependencies — that have a direct impact on the reproducibility of experiments — into three categories: infrastructural dependency, data dependency and job execution dependency. According to their work, reproducibility of computational studies requires to fully document the computational environments, and to ensure that all experimental resources remain accessible.

Chard et al. [18] highlight the importance of data publication systems in data-intensive science. The authors stress the need to describe requirements for data publishing and illustrate that sharing on simple and basic network-accessible storages — like a Dropbox folder — is insufficient. They demand published data to be identifiable, described, preserved and searchable, motivating the need for dedicated data publication systems.

Incentivizing Reproducible Practices. Missing rewards and incentive structures have been identified as core contributors to the reproducibility challenge. Studies highlight that conferences and journals may encourage or demand publishing relevant experiment data as part of the publication process [6, 54]. Other incentive structures are based on monetary benefits. Russell [50] demands funding agencies to reward scientists based on the reproducibility of their research. Rosenblatt [47] highlights the collaborative agreements between universities and the industry. Companies could provide financial benefits for reproducible data, thus improving the overall quality of the research collaboration. Understanding better the role of incentives in reproducible research practices will also be key in designing technology that supports reproducibility.

Replication in HCI. In HCI it is common to refer to *replication* of research. Wilson et al. [59] stress that novelty-driven research and diversity in HCI require discussing the place of replication in HCI. They describe four notions: *Direct replication* to validate findings; *Conceptual replication* refers to validity based on alternative approaches; *Replicate & Extend* means to reproduce prior research before making further investigations; and finally *Applied Case Studies* refers to application of research findings in real world contexts.

In their paper '*Is replication important for HCI?*', Greiffenhagen and Reeves [35] also stress the need to understand *aims and motivations* for replication in HCI. They argue to distinguish between *"what may be replicable and what is actually replic-ated."* While *replicable* means that research *in principal* can be replicated, *replic-ated* marks research that *has been replicated*. This distinction relates to the role of HCI in science, similar to *"psychology's own debates around its*

status as a science (that) are also consonant with these foundational concerns of 'being replicable'". The authors highlight that *"to focus the discussion of replication in HCI, it would be very helpful if one could gather more examples from different disciplines, from biology to physics, to see whether and how replications are valued in these."* In fact, as part of our study we aim to better understand the role and value of reproducibility in HEP. However, our study focuses on perceptions and design requirements for technology that supports reproducible research and is not designed to contribute *directly* to discussions on the role of replication in HCI.

Design for Supporting Research Practices

Research has shown that the design of scientific tools profits from taking a human-centered approach, instead of studying only technical requirements [42] and that even small changes to the interface of analysis systems leads to adapted behavior of scientists [37]. Given that impact, it is clear that successful service design requires involving domain experts [55] in the process. In fact, improving research infrastructures, e.g. for collaborative data generation and reuse, requires *"a deeper understanding of the social and technological circumstances"* [43], motivating our researcher-centered study approach.

In the context of research replicability, Mackay et al. [40] presented *Touchstone*, an experiment design platform for HCI research on interaction techniques. The authors highlight that it is difficult to compare new techniques to the variety of existing ones, because of the effort needed to replicate those. Thus, comparison is often done only for one standard technique. The described platform allows to specify experiments and supports researchers with the evaluation process. Experiment designs and log data can be exported and imported, enabling reuse, replication and extension of research.

As sharing of research enables accessibility and improves visibility, studies [44, 51] found a clear connection between citation benefits for publications and open sharing of their experiment data. Thus, concerning the design of a community data system, Garza et al. [31] found that emphasizing *"the potential of data citations can affect researchers' data sharing preferences from private to more open."* And also badges have proven to encourage research sharing. Kidwell et al. [38] compared contributions to the *Psychological Science* journal, that adopted open science badges, to other journals in the same domain that have not done so. Papers got a visible badge in case data or materials from the reported study were released, leading to a significant increase in data sharing. ACM introduced very similar and even more fine-grained open research badges that even promote rewarded publications in their digital library [1, 11].

3 RESEARCH CONTEXT

We conducted our study at the European Organization for Nuclear Research (CERN). The study profited from the amount of data recorded in CERN's experiments, the demonstrated early adoption of computer-supported technology and an existing, tailored research preservation service.

HEP, CERN and the LHC Collaborations

In recent years, CERN received attention for discoveries surrounding the Large Hadron Collider (LHC). The LHC is the world's largest and most powerful particle accelerator [26]. At four locations, particle collisions are measured by detectors, each of which is represented by a so-called LHC *collaboration*. The four main LHC collaborations are: ALICE, ATLAS, CMS and LHCb [36]. To be able to verify findings, LHC collaborations mostly perform their research independently from others. As Cho [20] highlights, that is especially true for CMS and ATLAS that have similar research goals, thus creating competition. Even though all research data are recorded locally within the detectors, LHC collaborations are not simply local organizational structures at CERN, but rather a global network that includes hundreds of institutes worldwide[1]. However, despite their global scale, CERN is their center point. Concerning the structure of LHC collaborations, Merali [41] argues that there is no simple top-down decision making, but rather a distribution of responsibility towards the many highly specialized teams. Merali further refers to a spokesperson who notes that "in industry, if people don't agree with you and refuse to carry out their tasks, they can be fired, but the same is not true in the LHC collaborations." That is because "physicists are often employed by universities, not by us." These are important aspects to consider in this study, as we can not rely on a central facilitator to command compliance with reproducible practices.

Despite competition between LHC collaborations, openness in scholarly communication is characteristic in HEP. The preprint server culture enables scientists to share ideas and results freely and immediately [23, 32]. In her ethnographic study, Velden [56] illustrates the openness that characterizes scholarly communication in HEP. She illustrates, how — despite competition — groups working with shared, large-scale facilities, share information in a relatively open fashion.

A pillar of the open research practices is the field's ability to develop and adapt to supportive technologies. It is not coincidental that the roots of the World Wide Web (WWW) lead back to CERN, where it was conceived to share data between institutes around the world [8, 9, 16]. And still today, HEP makes for a strong environment to study handling of

---

[1]https://greybook.cern.ch/greybook/researchProgram/detail?id=LHC

Figure 1: Part of the analysis submission form that allows physicists to describe and preserve their analyses. Supportive mechanisms ease efforts, ensure that data map to the internal LHC collaboration structures and guarantee consistency between records. In this scenario, researchers can chose between two possible types of datasets. Based on this choice, input in the following fields can be validated.

Interviewee reference	Affiliation	Gender	Experience
P1	ATLAS	Male	Postdoc
P2	LHCb	Male	PhD student
P3	LHCb	Male	Senior researcher
P4	CMS	Male	Postdoc
P5	CMS	Male	Postdoc
P6	CMS	Male	Senior researcher
P7	CMS	Male	Senior researcher
P8	CMS	Female	PhD student
P9	CMS	Male	Convener
P10	CMS	Male	Senior researcher
P11	LHCb	Male	Convener
P12	CMS	Male	PhD student

Table 1: Overview of the affiliations and professional experiences of the interviewees. We recruited data analysts from three LHC collaborations with a wide variety of experience. The male oversampling reflects the employment structure of research physicists at CERN.

unmatched data volumes, as HEP remains to be one of the most data-intensive branches of science [30].

CERN Analysis Preservation (CAP)

The CERN Analysis Preservation (CAP) prototype service[2] enables researchers from the LHC collaborations to describe their analyses, consisting of data, metadata, workflows and code files [19]. Stored descriptions, data and files are preserved. The service thereby supports key reproducibility requirements: rich data description and long-term preservation. One of the key elements of CAP is a web-based graphical user interface that allows physicists to easily describe their analyses. Figure 1 shows a part of the LHCb analysis submission form. Due to differences in data analysis structures, analysis preservation templates are tailored to the experiment to which they belong. Initially, analyses on CAP are accessible to the creator as drafts. They can be shared with the whole LHC collaboration or individual collaboration members. Analyses are not shareable between different LHC collaborations.

The prototype is currently tested in a joint effort with several LHC collaborations. It is designed as a service that provides an easy and consistent way of describing and storing LHC analyses. Efforts were taken to support researchers in the description process. Depending on the data that are stored in the individual collaboration databases, CAP tries to auto-complete and auto-suggest as much information as

possible. Nevertheless, the time required to fully describe and store an analysis is significant and adds to researchers' workload.

4 METHOD

We carried out 12 semi-structured interviews, to establish an empirical understanding of data sharing and preservation practices, as well as challenges and opportunities for systems that enable preservation and reproducibility.

Recruitment and Participants

In this section, we provide rich descriptions of the participants, including researchers' affiliations and experience levels. The analysts' ages ranged from 24 to 42 years old (average = 33, SD = 5.2). We decided not to provide information on the age of individual participants, as it would — in combination with the additional characteristics — allow to identify our participants. The 12 interviewees included 1 female (P8) and 11 males. The male oversampling reflects the employment structure at CERN: in 2017, between 79% and 90% (depending on the type of contract) of the research physicists working at CERN were male [17]. All interviewees were employed at CERN or at an institute collaborating with CERN. As all interviews were conducted during regular working hours, they became part of an analyst's regular work day. Accordingly, no additional remuneration was provided.

Collaborations and Experience. We interviewed data analysts working in three main LHC collaborations. Our recruitment focused on CMS and LHCb, as their preservation templates are most complex and developed. No interviewee had a hierarchical connection to any of the authors. Table 1 provides

---

[2]Publicly available on GitHub:
   https://github.com/cernanalysispreservation

an overview of the interviewees' affiliations with the LHC collaborations.

We selected physicists with a diverse level of experience and various roles to ensure a most complete representation of practices and perceptions. Half of the interviewees are early-stage researchers: PhD students and postdocs. The other half consists of senior researchers. As all interviewees - except the PhD students - held a PhD, we introduced metrics to distinguish between postdocs and senior researchers. In accordance with the maximum duration of postdoctoral fellowship contracts at CERN, we decided to consider as *senior researchers* all interviewees who had worked for more than three years as postdoctoral physics researchers.

Two of the senior researchers had a convening role, or had such responsibilities within the last two years. Conveners are in charge of a working group and have a project management view. They are, however, often working on analyses themselves. Since they have this unique role within LHC collaborations, we identified them separately in Table 1.

Cultural Diversity. According to 2017 personnel statistics [17], CERN had a total of 17,532 personnel, of which 3,440 were directly employed by the organization. CERN has 22 full member states, leading to a very diverse work environment. We decided not to list the nationalities of individual scientists, as several participants asked us not to do so and because we were concerned that participants could be identified based on the rich characterization already consisting of affiliation, experience and gender. However, we report the nationalities involved. The participants were in alphabetical order: British, Finnish, German, Indian, Iranian, Italian, Spanish and Swiss. The official working languages at CERN are English and French, with English being the predominant language in technical fields. All interviews were conducted in English. Working in a highly international environment at CERN, all interviewees had a full professional proficiency in English communication.

Interview Protocol

Initially, participants were invited to articulate questions and were asked to sign the consent form. The 12 interviews lasted on average 46 minutes (SD = 7.6). The semi-structured interviews followed the outline of the questionnaire:

Initially, questions targeted practices and experiences regarding analysis storage, sharing, access and reproducibility. Interviewees were encouraged to talk about expectations regarding a preservation service and the value of re-using analyses. This part of the questionnaire informed the themes MOTIVATION and COMMUNICATION. Next, we provided a short demonstration of the CAP prototype. Participants were introduced to the analysis description form and to collaborative aspects of the service: sharing an analysis with the LHC

collaboration and accessing shared work. Participants were asked to imagine the service as an operational tool and were invited to describe the kind of information they would want to search for.

We used two paper exercises to support the effort of uncovering the underlying structure of analyses, as perceived by data analysts. In one exercise, participants were asked to design a faceted search for a search result page, showing a set of analyses with abstract titles. They had three empty boxes at their disposal and could enter a title and four to seven characteristics each. In the second exercise, we encouraged participants to draw connections and dependencies that can exist between analyses on a printout with two circles, named *Analysis A* and *Analysis B*. The exercise supported us in understanding the value of a service being aware of relations between analyses. Finally, interviewees were encouraged to reflect on CAP and invited to describe how they keep aware of colleagues' ongoing analyses within their LHC collaboration.

The system-related part of the questionnaire and the paper exercises informed our results about UNCERTAINTY, COLLABORATION and AUTOMATION.

Data Analysis

All interviews were transcribed non-verbatim by the principal author. We used the Atlas.ti data analysis software to organize, code and analyze the transcriptions. Thematic analysis [10] was used to identify emerging themes from the interviews. We performed an initial analysis after the first six interviews were conducted. At first, we repeatedly read through the transcriptions and marked strong comments, problems and needs. Already at this stage, it became apparent that analysts were troubled by challenges the currently employed communication and analysis workflow practices posed. After we got a thorough understanding of the kind of information contained in the transcriptions, we conducted open coding of the first six interviews. As the principal author and two co-authors discussed those initial findings, we were content to see the potential our interviews revealed: the participants already described tangible examples of how a preservation service might motivate their contribution as a strategy to overcome previously mentioned challenges. We decided not to apply any changes to the questionnaire.

As the study evolved, we proceeded with our analysis approach and revised already existing codes. We aggregated them into a total of 34 code groups that were later revised and reduced to 22 groups. The reduction was mainly due to several groups describing different approaches of communication, learning and collaboration. For example, three smaller code groups that highlighted various aspects of e-mail communication were aggregated into one: *E-Mail (still) plays key role in communication.* We continued to discuss

our evolving analysis while conducting the remaining interviews. In addition, the transcript of the longest interview was independently coded by the principal author, one co-author and one external scientist, who gained expertise in thematic content analysis and was not directly involved in this study.

A late version of the paper draft was shared with the 12 interviewees and they were informed about their interviewee reference. We encouraged the participants to review the paper and to discuss any concerns with us. Eight interviewees responded (P2, P4, P5, P7, P8, P9, P11, P12), all of which explicitly approved of the paper. We did not receive critical comments regarding our work. P9 provided several suggestions, almost all of which we integrated. The CMS convener also proposed to *"argue that the under-representation of ATLAS is not a big issue, as it is likely that the attitudes in the two multi-purpose experiments are similar (the two experiments have the same goals, similar designs, and a similar number of scientists)."*

5 FINDINGS

Six themes emerged from our data analysis. In this section, we present each theme and our understanding of the constraints, opportunities and implications involved.

Motivation

Our analysis revealed that personal motivation is a major concern in research preservation practices. In particular P1, P2, P7, P9 and P11 worry about contribution behaviors towards a preservation service. P1 further contrasts information *use* and *contribution*: *"People may want to use information - but we need to get them to contribute information as well."* The analyst calls this *"the most difficult task"* to be accomplished.

Several analysts (P1, P2, P9, P11) point to missing incentives as the core challenge. They stress that preserving data is not immediately rewarding for oneself, while requiring substantial time and effort. P9 highlights that even though analysts who preserve and share their work might get slightly more citations, this is *"a mild incentive. It's more motivating to start a new analysis, other than spending time encoding things..."*

In this context, convener P11 critically contrasted policies with resulting preservation quality and highlighted the motivational strength of returned benefits:

"...if you take this extra step of enforcing all these things at this level, it's never going to get done. Because if you use this as a documentation, so I'm done, now I'm going to put these things up. If it complains, like, I don't care... [...] But if there is a way of getting an extra benefit out of this, while doing your proper preservation, that is good - that would totally work."

Imagining a service that not only provides access to preserved resources, but allows systematic execution of those,

the convener states that he does not *"see any attitude problem anymore, because doing this sort of preservation gives you an advantage."* Such immediate mechanisms might also provide incentives to integrate a preservation service into the analysis workflow, which according to P9 will be crucial. The convener expects that researchers *"will not adapt to data preservation afterwards. Or five percent will do."*

Communication

Our analysis revealed that data analysts in HEP have a high demand for information. Yet, communication practices often depend on personal relations. All of our interviewees described the need to access code files from colleagues or highlighted how access could support them in their analysis work. Even though most analysts (P2 - P4, P6 - P8, P10 - P12) explicitly stated that they share their work on repositories that provide access to their LHC collaboration, information and resource flow commonly relied on traditional methods of communication:

"The few times that I have used other people's code, I think that...I think it was sent to me by e-mail all the times" (P3)

"They have saved their work and then I can ask them: 'where have you located this code? Can I use it?' And they might send me a link to their repository." (P8)

The analysis of our interviews revealed the general practice of engaging in personal communication with colleagues in order to find resources. P4 highlights a common statement, i.e. colleagues pointing to existing resources:

"You go to the person you know is working on that part and you ask directly: 'Sorry, do you know where I can find the instructions to do that?' and he will probably point to the correct TWiki or the correct information"

Personal relations are vital in this communication and information architecture. Most analysts (P1, P2, P3, P4, P6, P7, P8, P9, P11) stressed that it was important to know the right people to ask for information. P8 described the effort needed:

"I mean you have to know the right people. You have to know the person who maybe was involved in 2009 in some project. And then you have to know his friend, who was doing this. And his friend and then there is somebody who did this and she can tell you how it went."

But, communication and information exchange was often contained within groups and institutes. P7 stressed that for a certain technique, other groups *"have better ideas. In fact, I know that they have better ideas than other groups, but they are not using them, because we are not talking to each other."* P2 stated that *"being shy and not necessarily knowing who to e-mail"* are personal reasons not to engage in communication with colleagues. The challenge to find the right colleagues to
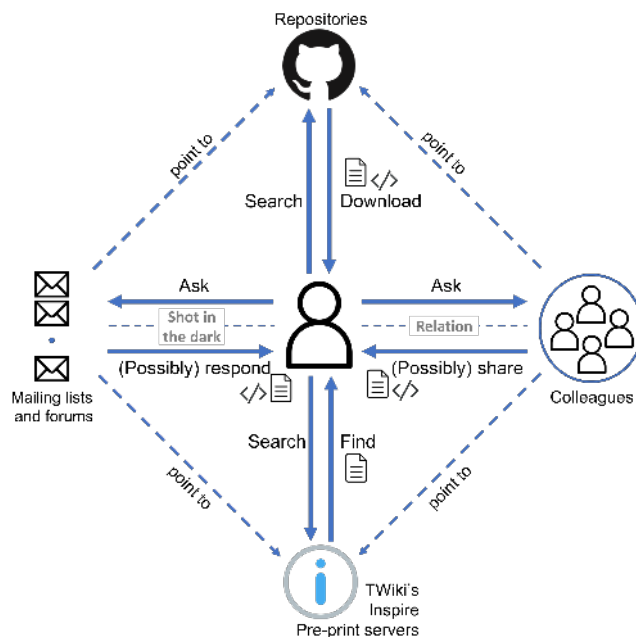
Figure 2: The current flow of information in HEP data analysis is characterized by the need to ask colleagues and the uncertainty of finding required resources.

talk to is increased by the high rotation of researchers, many of them staying only few years.

Almost all analysts (P1 — P4, P6 — P11) in our study referred to another common issue they encounter: the lack of documentation. P6 illustrated the link between missing documentation and the need to ask for information instead:

"This is really mouth-to-mouth how to do this and how to do that. I mean the problem for preservation is that at the moment it's just: ask your colleague, rather than write a documentation and then say 'please read this.'"

Meetings and presentations are a key medium in sharing knowledge. However, the practice of considering presentations as a form of knowledge documentation makes access to information difficult:

"There are cases you asked somebody: 'but did they do this, actually?' And somebody says like: 'I remember! Two years ago, there was this one summer meeting. We were having coffee and then they showed one slide that showed the thing.' And this slide might have never made it to the article." (P8)

Uncertainty

Our interview findings revealed that the communication and information architecture leads to two types of *uncertainty*:

(1) related to the *accessibility* of information and resources; and (2) connected to the *volatility* of data.

Accessibility. As depicted in Figure 2, analysts follow two principal approaches to *access* information and resources: they search for them on repositories and databases or ask colleagues. The outcome of directly searching for resources contains uncertainty, as researchers might not be sure exactly what and where to search. But, also various search mechanisms represent challenges. A researcher described searching for an analysis and highlights, that *"at the moment, it's sometimes hard to find even the ones that I do know exist, because I don't know whether or not they are listed maybe under the person I know. So, [name] I know that I can find... Well, actually I don't know if I can find his analysis under his GitHub user." (P2)*

Our interviewees (P1 — P4, P6 — P9, P11, P12) reported that they typically contact colleagues or disseminate requests on mailing lists and forums to ask for information and resources. While mailing lists represent a shot in the dark, the success of approaching colleagues is influenced by personal relations. If successful, they receive required resources directly or are pointed to the corresponding location.

Volatility. Facing vast amounts of data and dependencies, analysts wish that a centralized preservation service helps them with uncertainty that is caused by the *volatility* of data.

Analysis Integrity: A service aware of analysis dependencies can ensure that needed resources are not deleted.

"...and this can be useful even while doing the analysis, because what happens is that people need to make disk space and then they say: 'ah, we want to remove this and this and this dataset - if you need it, please complain.' And if you had this in a database for example, it could be used also saying like 'ah, this person is using this for this analysis' even before you would share your analysis." (P6)

The analyst even highlighted the possibility to track datasets of work in progress that was not yet shared with the LHC collaboration. A convener also motivates the issue that comes with the removal of data and describes the effort and uncertainty involved in current communication practices:

"Sometimes versions get removed from disk [...] And the physics planning group asks the conveners: 'ok, is anybody still using those data?' [...] I have to send an email of which version they are using etc. [...] And at some point, if I have 30 or 40 analyses going on in my working group, it's very hard not to make a mistake in this sense if people don't answer the emails. While if I go here, I say ok, this is the data they are using - I know what they are using - and it takes me ten minutes and I can have a look and I know exactly." (P11)

Receiving vital analysis information: We learned that different analyses often have input datasets in common. When

an analyst finds issues with a dataset, she or he draws back to the existing communication architecture.

"I present it in either one of the meetings which is to do with like that area of the detector for example. Or if it was something higher profile than maybe one of the three or four meetings which are more general, applicable to the collaboration[3]*. And from there that would involve talking to enough people in the management and various roles...that it would then I guess propagate to...they would be again in touch with whoever they knew about that might be affected." (P2)*

The risk of relying on this communication flow is that one might naturally miss vital information. An analyst could be unavailable to attend the right meeting or generally not be part of it. The person sending the email might also not know about all affected analyses. This might especially be true for relevant analyses that are conducted in a working group different from the ones of the analysts that are signaling the issue. A preservation service enabling researchers to signal warnings associated with a dataset or, generally, resources that are shared by various analyses, allows informing dependent analysts in a reliable manner. As being informed about discovered issues can be vital for researchers, it would be in their very interest to keep their ongoing analyses well documented in the service.

Staying Up-to-Date: Keeping up-to-date on relevant changes can be challenging in data-intensive environments. Researchers hope that a preservation service provides reliable dependency awareness to analysts who document their work:

"The system probably tells me: 'This result is outdated. The input has changed'. Technical example. At the moment, this communication happens over email essentially" (P6)

P11 told us about a concrete experience:

"He was using some number, but then at some point the new result came out and he had not realized. Nobody realized. And then, of course, when he went and presented things he was very advanced, they said 'well, there is a new result - have you used this? No, I have not used it.'"

Collaboration

Sharing their work openly, analysts increase their chance to engage in collaboration. Currently, useful collaboration is hindered by missing awareness of what others do. We can imagine this to be especially true outside of groups and dislocated institutes. P4 emphasizes the value of collaboration:

"The nTuple production is a really time consuming part of the analysis. So, if we can produce one set of nTuples...so one group produces them and then they can be shared by many analysis teams...this has, of course, a lot of benefits."

---

[3]The interviewee is referring to the LHC collaboration.

Researchers who document their ongoing activities and interests increase their discoverability within the LHC collaboration. Thereby, they increase their chance to be asked to join an official request that might satisfy their data needs:

"I want to request more simulation. [...] I would search and I would say these are the people. I would just write to them, because I want to do this few modifications. But maybe this simulation is also useful for them, so we can just get together and get something out." (P11)

In fact, a convener stated that due to the size of LHC collaborations, it is difficult to be aware of other ongoing analyses:

"CMS is so big that I cannot know if someone else is already working on it. So, if this tool is intended to have also the ongoing analyses since a very early stage, this would help me if I can know who is working on that." (P9)

P8 highlights that being aware of other analyses can possibly lead to collaboration and prevent unwanted competition:

"Because the issue at CMS - and probably at whole CERN - is that you want start working on it, but, on the other hand, it's rude if you start working on something and you publish and then you get an angry message, saying: 'hey, we were just about to publish this, and you cannot do it.' [...] The rule is that everyone can study everything, but, of course, you don't want to steal anybody's subjects. So, if it wouldn't be published, you would then maybe collaborate with them."

Automation

We see an opportunity to support researchers based on the common structure that applies to analyses: *"because in the end, everybody does the same thing"* (P7). A convener characterized this theme by demanding *"more and more Lego block kind analyses, keeping to a minimum the cases where you have to tailor the analysis a bit out of the path"* (P9).

Templated analysis design. As P11 articulates, the common steps and well-defined analysis structure represent an opportunity to provide checklists and templates that facilitate analysis work:

"If, of course, I have some sort of checklist or some sort of template to say 'what is your bookkeeping queries — use this and that', then of course this would make my life easier. Because I would be sure I don't forget anything."

The convener makes two claims on how a structured analysis description template could support researchers. First, templates help in the analysis design. Second, the service could inform about missing fragments or display warnings based on a set of defined checks. However, it is important to recognize a core challenge that comes with well-structured analysis templates; allowing for sufficient flexibility:

"Somehow these platforms tend to — which is one of the strong points, but at the same time one of the weaknesses — is that

[...] it gives you some sort of template and makes it very easy for you to fill in the blanks. But at the same time, this makes things difficult, if you want to make very complex analyses where it's not so obvious anymore what you want to do." (P11)

Automate Running and Interpretation. Several analysts (P2, P5, P7, P8, P11) expressed their wish for centralized platforms to automate tasks that they would currently have to perform manually. P2 stated:

"So, being able to kind of see that it...might be able to submit to it and then it just goes through and runs and does everything...and I don't need to think too much about whether or not something is going to break in the middle for something that is nothing related to me, would potentially be quite nice."

However, not only automating the full execution of analyses seems desirable, but also interpretation of systematics:

"And I say: 'ok, now I want to know for example, which are the systematics' and you can tell me, because you know you have the information to do it by yourself. You will save a lot of time. People will be very happy I think." (P5)

Preventing mistakes. P7 described how the similarity and common structure of analyses supports automated comparison and verification:

"What I would like to search is the names of the Monte Carlo samples used by other analyses. [...] the biggest mistake you can make is to forget one. Because if you forgot one, then you will see new physics, essentially. And it's a one-line mistake."

Developing a feature that compares a list of dataset identifiers and that points to irregularities is trivial. Yet, as P7 continues to describe the effort needed to do the comparison at the moment, the perceived gain seems to be high:

"So, the analysis note always contains a table - it's a PDF. Then always contains a table with a list of Monte Carlos. I often download that, look at the table and see what's missing. Copy paste things from there. But so here, I would be able to do it directly here."

Scalability

Although not directly in the scope of the questionnaire, four interviewees (P3, P8, P9, P11) commented on the growing complexity of analysis work in HEP, stressing the importance of preservation and reproducibility. P9 highlights the issues that evolve from collecting more and more data:

"As we collect the data, the possibility of analysis grows. In fact, we are more and more understaffed, despite of being so many in the collaboration[4]. Because, what is interesting for the particle physics community grows as data grow. And so, we get thinner and thinner in person power in all areas that we deem crucial."

[4]The interviewee is referring to the LHC collaboration.

The convener adds that *"a typical analysis cycle becomes much much longer. Typical contract duration stays the same."* P3 details how the high amount of rotation and (ir-)reproducibility impact analysis durations:

"If someone goes and an analysis is not finished, it might take years. Because there was something only this person could do. I think that analysis preservation could help a lot on this. [...] But otherwise you might have to study analyses from scratch if someone important disappears."

P11 agrees that *"it's getting more and more complex, so I think you really need to put things together in a way that is reasonable and re-runnable in some sort of way."* P9 coined the term *orphan analyses*. It describes analyses for which no one is responsible anymore. The convener expects that *"at some point it will become a crisis. Because, so far, it was a minority of cases of orphan analyses. It will become more and more frequent, unless contract durations will change. But this will not happen."*

6 IMPLICATIONS FOR DESIGN

We present challenges and opportunities in designing for research preservation and reproducibility. Our work shows that the ability to access documented and shared analyses can profit both individual researchers and groups [27]. Our findings hint towards what Rule et al. [49] call *"tension between exploration and explanation in constructing and sharing"* computational resources. Here, we primarily learned about the **need to motivate and incentivize contributions**. Based on our findings, we show how design can create motivating, *secondary usage forms* of the platform and its content, related to uncertainty, collaboration and structure. And, while references in this section underline that the CHI community has established a long tradition of studying collaboration and communication around knowledge work, it is not yet known how to design collaborative systems that foster reproducible practices and incentivize preservation and data sharing. The following description of *secondary usage forms* aims to contribute to knowledge about motivations and incentives for platforms that support research reproducibility.

Exploit Platforms' Secondary Functions

As observed in the MOTIVATION theme, getting researchers to document and preserve their work is a main concern. In this context, researchers critically commented on the impact of policies, creating little motivation to ensure the preservation quality beyond fulfilling formal requirements. And also citation benefits, commonly discussed as means to encourage research sharing [44], might provide only a mild incentive, as time required for documenting and preserving can be spend more rewarding on novel research. This seems especially true in view of growing opportunities that result

from the increasing amount of data, as described in the Scal-ability theme. Yet, researchers indicated how centralized preservation technology can uniquely benefit their work, in turn creating motivation to contribute their research. Thus, we have to **study researchers' practices, needs and challenges in order to understand how scientists can benefit from centralized preservation technology. Doing so, we learn about the** *secondary function* **of the platform and its content, crucial in developing powerful incentive structures**.

Support Coping with Uncertainty

As we learned in the Communication theme, the information architecture is heavily relying on personal connections and communication, leading to a high degree of Uncertainty related to the *accessibility* and *volatility* of information and data. Consequently, researchers report encountering severe issues related to the insufficient transparency and structure that a centralized preservation service might be able to mitigate. We propose two strategies: First, a centralized preservation service can implement overviews and details of analysis dependencies not available anywhere else. Implementing corresponding features enables us to **promote preservation as effective strategy to cope with uncertainty** so that research integrity of documented dependencies can be guaranteed. Second, we further imagine documenting analyses on a dedicated, centralized service to be a powerful strategy to *minimize* uncertainty towards updated dependencies and erroneous data, if the service provides awareness to researchers. In the case of data-related warnings, reliable notifications could be sent to analysts who depend on collaboration-wide resources, replacing current, less reliable communication architectures. This approach also relates to uncertainties at the *data layer*, as described by Boukhelifa et al. [14], who studied types of uncertainty and coping strategies of data workers in various domains. According to their work, the three main active coping strategies are: *Ignore*, *Understand* and *Minimize*. In summary, our findings suggest that such secondary benefits might drive researchers to contribute and use the preservation tool.

Provide Collaboration-Stimulating Mechanisms

The Collaboration theme highlighted the importance of cooperation in HEP. Analysts save time when they join forces with colleagues or groups with similar interests. Yet, awareness constraints resulting from the communication and information architecture often hinder further collaboration. We postulate that the preservation platform can add useful secondary benefits for theses cases. First, given the centralized interface and knowledge aggregation function of a preservation service, we see opportunities to **support locating expertise in research collaborations**. In

fact, especially knowledge-intensive work profits from such supporting tools, as it enables sharing expertise across organizational and physical barriers [22]. Ehrlich et al. [25] note that awareness of "who knows what" is indeed key to stimulating collaboration. In an organizational context, Transactive Memory Systems (TMS) are employed to create such awareness. HEP collaborations are TMS, in that the sum of knowledge is distributed among their analysts and the communication between them forms a group memory system [57]. Further research on the support and integration of TMS in the context of platforms for research reproducibility could increase acceptance through heightened awareness provided by such platforms. Also, elements of social file sharing could further stimulate discovery and exploration of relevant researchers and analyses. As noted by Shami et al. [52], this can be particularly important in large organizations.

Second, an important benefit could be the visibility of team or project members. Taking preserved research as basis for expertise location can incentivize contributions, as scientists who document in great detail are naturally most visible, thus increasing their chances to engage in collaboration. This approach also enables us to mitigate privacy concerns, by considering only resources of analyses that have been shared with the LHC collaboration. Mining documented and shared research to provide expertise location thus mitigates common challenges: Typically, workplace expertise locators infer knowledge either by mining existing organizational resources like work emails [15, 34], or by asking employees to indicate their skills and connections within an organization [53]. While automated mining of resources may cause privacy concerns, relying on users to undergo the effort of maintaining an accurate profile is slower and less complete [46]. Given the increased interdisciplinary and international research culture, developing such bridging mechanisms — even though not central to the service missions — is especially helpful.

Support Structured Designs

A community-tailored research preservation service can support analysts through automated mechanisms that make use of prevalent workflow structures. Researchers pointed out that analysis work within a LHC collaboration commonly follows general patterns, demanding even to further streamline processes as much as possible; thereby pointing to the guiding role of preservation technology. We propose to **design community-tailored services that closely map research workflows to preservation templates**. That way, preservation services can provide checklists and guidance for the research and preservation process; furthermore, automation of common workflow steps can increase efficiency. Additionally, if the preservation service is well embedded

into the research workflows, it could enable supportive mechanisms like auto-suggest and auto-completion. Such steps are key to minimizing the burden of research preservation, which is of great importance, as we acknowledge that the acceptance and willingness to comply with reproducible practices will always be related to the cost/benefit ratio related to research preservation and sharing. Having noted the need for automation and taylorization of interfaces, we need to emphasize the significance of academic freedom when designing such services. Design has to account for all the analyses, also those that are not reflected in mainstream workflows. We have to **support creativity and novelty by leaving contributors in control**. This applies both for supportive mechanisms like auto-complete and auto-suggest, as well as for the template design.

7 DISCUSSION

The study's findings and implications have pointed to several relationships that are important for designing technology that enables research preservation and reproducibility. First, we have contrasted required efforts with returned benefits. It is apparent that stimuli are required to encourage researchers to conduct uninteresting and repetitive documentation and preservation tasks that in itself, and at least in the short run, are mostly unrewarding. Thus, not surprisingly, the call for policies is prominent in discussions on reproducible research. Yet, our findings hint towards the relation between preservation quality and policies, raising doubts that policies can encourage sustained commitment to documentation and preservation beyond a formal check of requirements. In this context, we hypothesize that also the relation between policies and flexibility needs to be considered. Thinking about structured description mechanisms as provided by CAP, one needs to decide on a common denominator that defines the main building blocks to comply with the policies. However, this is likely to create two problems: (1) Lack of motivation to preserve fragments that are not part of the basic building blocks of research conducted within the hierarchical structure for which the policies apply; (2) Preservation platforms that map policies might discourage or neglect research that is not part of the fundamental building blocks.

Facing those conflicting relationships, meaningful incentive structures could positively influence the reproducibility challenge and create a favorable shift of balance between required efforts and returned benefits. We postulate that communities dealing with the design of such systems need to invest a significant amount of time into user research to create tailored and structured designs. Further research in this area is surely needed, i.e. the evaluation of prototypes or established systems in general and with a focus on the users' exploitation of secondary benefits of the system. This call for more research in this area is particularly evident when

looking at the latest study by Rowhani-Farid et al. [48] who found only one evidence-based incentive for data sharing in their systematic literature review. They conducted their study in search of incentives in the health and medical research domain, one of the branches of science that was in the focus of reproducibility discussions from the very beginning. The only reported incentive they found relates to open science badges that resulted in a significant impact in data sharing of papers submitted to the *Psychological Science* journal. The authors highlight that *"given that data is the foundation of evidence-based health and medical research, it is paradoxical that there is only one evidence-based incentive to promote data sharing. More well-designed studies are needed in order to increase the currently low rates of data sharing."*

Our study showed how design can create secondary usage forms of preservation technology and its content related to communication, uncertainty, collaboration and automation. Described mechanisms and benefits apply not only to submissions at the end of the research lifecycle, but, rather, provide certainty and visibility for ongoing research. The significance of such contribution-stimulating mechanisms is particularly reflected in the observed scalability challenge, indicating that reproducibility in data-intensive computational science is not only a scientific ideal, but a hard requirement. This is particularly notable as the barriers to improve reproducibility through sharing of digital artefacts are rather low. Yet, it must also be noted that not all software and data can always be freely and immediately shared. The claim for reproducibility does not overrule any legal or privacy concerns. Our results apply primarily to datasets generated through experiments without human participants. Future research should investigate incentives and requirements for sharing data from human subject research.

8 LIMITATIONS AND FUTURE WORK

We aim to foster the reproducibility of our work and to provide a base for future research. Therefore, this paper is accompanied by various resources from our study. Those include the semi-structured interview questionnaire, the ATLAS.ti code group report and the templates of the two paper exercises. As is the core idea of reproducible research, we envision future work to extend and enrich our findings and design implications by studying perceptions, opportunities and challenges in diverse scientific fields. We can particularly profit from empirical findings in fields that are characterized by distinct scholarly communication and field practices and a differing role of reproducibility. Also different forms of research will need to be studied. Our study's focus is on data-intensive natural science, using the example of computational research in HEP. It does not intend to contribute directly to other forms of research such as descriptive and qualitative research.

It should also be noted as a limitation of the study that the reference preservation service is based entirely on custom templates. While this does not reflect the majority of repositories and cloud services used today for sharing research, our findings indicate that templates are key to enable and support secondary usage forms. And even though our study focused solely on HEP, findings and implications are however likely to be relevant for numerous fields, in particular computational and data-driven ones. Uncertainty, visibility and automation are of general concern to researchers, with HEP representing an ideal study context that provides one of the most data-intensive, diverse, distributed and technology-adopting environments.

9 CONCLUSION

This paper presented a systematic study of perceptions, opportunities and challenges involved in designing technology that enables research preservation and reproducibility in High Energy Physics, one of the most data-intensive branches of science. The findings from our interview study with 12 experimental physicists highlight the resistance and missing motivation to preserve and share research, core requirements of reproducible science. Given that the effort needed to follow reproducible practices can be spent on novel research — usually perceived to be more rewarding — we found that contributions to research preservation technology can be stimulated through secondary benefits. Our data analysis revealed that contributions to a centralized preservation platform can target issues and improve efficiency related to COMMUNICATION, UNCERTAINTY, COLLABORATION and AUTOMATION. Based on these findings, we presented implications for designing technology that supports reproducible research. First, we discussed how studying researchers' practices enables exploiting secondary usage forms of platforms and its content that are expected to stimulate researchers' contributions. Centralized repositories can promote preservation as an effective strategy to cope with uncertainty; support locating expertise in research collaboration; and provide a more guided and efficient research process through preservation templates that closely map research workflows.

ACKNOWLEDGMENTS

REFERENCES

[1] ACM. 2018. Artifact Review and Badging. Website. (April 2018). https://www.acm.org/publications/policies/artifact-review-badging Retrieved September 10, 2018.

[2] Monya Baker. 2016. 1,500 scientists lift the lid on reproducibility. *Nature* 533, 7604 (2016), 452–454. https://doi.org/10.1038/533452a

[3] Anna Bánáti, Péter Kacsuk, and Miklós Kozlovszky. 2015. Four level provenance support to achieve portable reproducibility of scientific workflows. *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2015 - Proceedings* (2015), 241–244. https://doi.org/10.1109/MIPRO.2015.7160272

[4] Sean Bechhofer, Iain Buchan, David De Roure, Paolo Missier, John Ainsworth, Jiten Bhagat, Philip Couch, Don Cruickshank, Mark Delderfield, Ian Dunlop, Matthew Gamble, Danius Michaelides, Stuart Owen, David Newman, Shoaib Sufi, and Carole Goble. 2013. Why linked data is not enough for scientists. *Future Generation Computer Systems* 29, 2 (2013), 599–611. https://doi.org/10.1016/j.future.2011.08.004

[5] C. Glenn Begley and Lee M. Ellis. 2012. Drug development: Raise standards for preclinical cancer research. *Nature* 483, 7391 (2012), 531–3. https://doi.org/10.1038/483531a arXiv:arXiv:cond-mat/9907372v1

[6] Khalid Belhajjame, Jun Zhao, Daniel Garijo, Kristina Hettne, Raul Palma, Óscar Corcho, José-Manuel Gómez-Pérez, Sean Bechhofer, Graham Klyne, and Carole Goble. 2014. The Research Object Suite of Ontologies: Sharing and Exchanging Research Data and Methods on the Open Web. *arXiv preprint arXiv: 1401.4307* February 2014 (2014), 20. arXiv:1401.4307 http://arxiv.org/abs/1401.4307

[7] Gordon Bell, Jim Gray, and Alex Szalay. 2006. Petascale Computational Systems: Balanced Cyber-Infrastructure in a Data-Centric World. *IEEE Computer* 39 (2006), 110–113. https://doi.org/10.1109/MC.2006.29

[8] Richard Bentley, Thilo Horstmann, Klaas Sikkel, and Jonathan Trevor. 1995. Supporting collaborative information sharing with the World Wide Web: The BSCW shared workspace system. In *Proceedings of the 4th International WWW Conference*, Vol. 1. 63–74.

[9] Tim Berners-Lee, Robert Cailliau, Jean-Francois Groff, and Bernd Pollermann. 1992. World-wide web: The information universe. , 52–58 pages. http://links.emeraldinsight.com/doi/10.1108/eb047254

[10] Ann Blandford, Dominic Furniss, and Stephann Makri. 2016. *Qualitative HCI Research: Going Behind the Scenes*. Morgan & Claypool Publishers, 51–60. https://doi.org/10.2200/S00706ED1V01Y201602HCI034

[11] Ronald F Boisvert. 2016. Incentivizing reproducibility. *Commun. ACM* 59, 10 (2016), 5–5.

[12] Philippe Bonnet, Stefan Manegold, and Matias Bjørling. 2011. Repeatability and workability evaluation of SIGMOD 2011. *ACM SIGMOD . . .* (2011), 45–48. https://doi.org/10.1145/2034863.2034873

[13] Christine L Borgman. 2007. *Scholarship in the digital age: information, infrastructure, and the internet*. MIT Press, Cambridge, MA.

[14] Nadia Boukhelifa, Marc-Emmanuel Perrin, Samuel Huron, and James Eagan. 2017. How Data Workers Cope with Uncertainty: A Task Characterisation Study. *Chi '17* (2017), 3645–3656. https://doi.org/10.1145/3025453.3025738

[15] Christopher S. Campbell, Paul P. Maglio, Alex Cozzi, and Byron Dom. 2003. Expertise identification using email communications. *Cikm 2003* January (2003), 528–531. https://doi.org/10.1145/956863.956965

[16] CERN. 2013. The birth of the web. Website. (Dec 2013). http://cds.cern.ch/record/1998446 Retrieved March 15, 2018.

[17] CERN. 2017. CERN Annual Personnel Statistics 2017. (2017). https://cds.cern.ch/record/2317058 Personnel Statistics 2017.

[18] Kyle Chard, Jim Pruyne, Ben Blaiszik, Rachana Ananthakrishnan, Steven Tuecke, and Ian Foster. 2015. Globus data publication as a service: Lowering barriers to reproducible science. In *Proceedings - 11th IEEE International Conference on eScience, eScience 2015*. 401–410. https://doi.org/10.1109/eScience.2015.68

[19] Xiaoli Chen, Sünje Dallmeier-Tiessen, Anxhela Dani, Robin Dasler, Javier Delgado Fernández, Pamfilos Fokianos, Patricia Herterich, and Tibor Šimko. 2016. CERN Analysis Preservation: A Novel Digital Library Service to Enable Reusable and Reproducible Research. In *International Conference on Theory and Practice of Digital Libraries*. Springer, 347–356.

[20] Adrian Cho. 2011. Particle Physicists' New Extreme Teams. *Science* 333, 6049 (2011), 1564–1567. https://doi.org/10.1126/science.333.6049.1564 arXiv:http://science.sciencemag.org/content/333/6049/1564.full.pdf

[21] Open Science Collaboration. 2012. An Open, Large-Scale, Collaborative Effort to Estimate the Reproducibility of Psychological Science. *Perspectives on Psychological Science* 7, 6 (2012), 657–660. https://doi.org/10.1177/1745691612462588

[22] Rob Cross and Jonathon N. Cummings. 2004. Tie and network correlates of individual performance in knowledge-intensive work. *Academy of Management Journal* 47, 6 (2004), 928–937. https://doi.org/10.2307/20159632

[23] Alessandro Delfanti. 2016. Beams of particles and papers: How digital preprint archives shape authorship and credit. *Social Studies of Science* 46, 4 (2016), 629–645. https://doi.org/10.1177/0306312716659373 arXiv:http://dx.doi.org/10.1177/0306312716659373

[24] Chris Drummond. 2009. Replicability is not reproducibility: Nor is it good science. *Proceedings of the Evaluation Methods for Machine Learning Workshop 26th International Conference for Machine Learning* 2005 (2009), 1–4.

[25] Kate Ehrlich, Ching-Yung Lin, and Vicky Griffiths-Fisher. 2007. Searching for experts in the enterprise: combining text and social network analysis. In *Proceedings of the 2007 international ACM conference on Supporting group work*. ACM, 117–126.

[26] Lyndon Evans and Philip Bryant. 2008. LHC Machine. *Journal of Instrumentation* 3, 08 (2008), S08001. https://doi.org/10.1088/1748-0221/3/08/S08001

[27] Davide Falessi, Giovanni Cantone, and Martin Becker. 2006. Documenting design decision rationale to improve individual and team design decision making: an experimental evaluation. In *Proceedings of the 2006 ACM/IEEE international symposium on Empirical software engineering*. ACM, 134–143.

[28] Dror G. Feitelson. 2015. From Repeatability to Reproducibility and Corroboration. *ACM SIGOPS Operating Systems Review* 49, 1 (2015), 3–11. https://doi.org/10.1145/2723872.2723875

[29] FORCE11. 2014. The FAIR data principles. Website. Retrieved August 8, 2017 from https://www.force11.org/group/fairgroup/fairprinciples.

[30] Mélissa Gaillard and Stefania Pandolfi. 2017. CERN Data Centre passes the 200-petabyte milestone. (Jul 2017). http://cds.cern.ch/record/2276551

[31] Kristian Garza, Carole Goble, John Brooke, and Caroline Jay. 2015. Framing the Community Data System Interface. In *Proceedings of the 2015 British HCI Conference (British HCI '15)*. ACM, New York, NY, USA, 269–270. https://doi.org/10.1145/2783446.2783605

[32] Anne Gentil-Beccot, Salvatore Mele, and Travis C. Brooks. 2010. Citing and reading behaviours in high-energy physics. *Scientometrics* 84, 2 (2010), 345–355. https://doi.org/10.1007/s11192-009-0111-1 arXiv:0906.5418

[33] Omar S. Gómez, Natalia Juristo, and Sira Vegas. 2010. Replications types in experimental disciplines. In *Proceedings of the 2010 ACM-IEEE International Symposium on Empirical Software Engineering and Measurement - ESEM '10*. 1. https://doi.org/10.1145/1852786.1852790

[34] Gopakumar Gopalakrishnan, Krupa Benhur, Abhishek Kaushik, and Anjaneyulu Passala. 2017. Professional Network Analytics Platform for Enterprise Collaboration. *Cscw 2017* (2017), 5–8. https://doi.org/10.1145/3022198.3023264

[35] Christian Greiffenhagen and Stuart Reeves. 2013. Is Replication important for HCI? *CEUR Workshop Proceedings* 976 (2013), 8–13.

[36] Hans Ake Gustafsson. 2006. LHC experiments. *Nuclear Physics A* 774, 1-4 (2006), 361–368. https://doi.org/10.1016/j.nuclphysa.2006.06.056

[37] Radu Jianu and David Laidlaw. 2012. An Evaluation of How Small User Interface Changes Can Improve Scientists' Analytic Strategies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 2953–2962. https://doi.org/10.1145/2207676.2208704

[38] Mallory C. Kidwell, Ljiljana B. Lazarević, Erica Baranski, Tom E. Hardwicke, Sarah Piechowski, Lina Sophia Falkenberg, Curtis Kennett, Agnieszka Slowik, Carina Sonnleitner, Chelsey Hess-Holden, Timothy M. Errington, Susann Fiedler, and Brian A. Nosek. 2016. Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing Transparency. *PLoS Biology* (2016). https://doi.org/10.1371/journal.pbio.1002456

[39] Jeffrey T Leek and Roger D Peng. 2015. Opinion: Reproducible research can still be wrong: adopting a prevention approach. *Proceedings of the National Academy of Sciences of the United States of America* 112, 6 (2015), 1645–6. https://doi.org/10.1073/pnas.1421412111 arXiv:1502.03169

[40] Wendy E Mackay, Caroline Appert, Michel Beaudouin-Lafon, Olivier Chapuis, Yangzhou Du, Jean-Daniel Fekete, and Yves Guiard. 2007. Touchstone: exploratory design of experiments. *CHI '07 Proceedings of the SIGCHI Conference on Human Factors in Computing System* (2007), 1425–1434. https://doi.org/10.1145/1240624.1240840

[41] Zeeya Merali. 2010. The large human collider: social scientists have embedded themselves at CERN to study the world's biggest research collaboration. Zeeya Merali reports on a 10,000-person physics project. *Nature* 464, 7288 (2010), 482–485.

[42] Jesper Molin, Paweł W. Woźniak, Claes Lundström, Darren Treanor, and Morten Fjeld. 2016. Understanding Design for Automated Image Analysis in Digital Pathology. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction (NordiCHI '16)*. ACM, New York, NY, USA, Article 58, 10 pages. https://doi.org/10.1145/2971485.2971561

[43] Gerard Oleksik, Natasa Milic-Frayling, and Rachel Jones. 2012. Beyond Data Sharing: Artifact Ecology of a Collaborative Nanophotonics Research Centre. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (CSCW '12)*. ACM, New York, NY, USA, 1165–1174. https://doi.org/10.1145/2145204.2145376

[44] Heather A. Piwowar and Todd J. Vision. 2013. Data reuse and the open data citation advantage. *PeerJ* 1 (Oct. 2013), e175. https://doi.org/10.7717/peerj.175

[45] Florian Prinz, Thomas Schlange, and Khusru Asadullah. 2011. Believe it or not: how much can we rely on published data on potential drug targets? *Nature reviews. Drug discovery* 10, 9 (2011), 712. https://doi.org/10.1038/nrd3439-c1 arXiv:arXiv:cond-mat/9907372v1

[46] Tim Reichling and Volker Wulf. 2009. Expert Recommender Systems in Practice : Evaluating Semi-automatic Profile Generation. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2009), 59–68. https://doi.org/10.1145/1518701.1518712

[47] Michael Rosenblatt. 2016. An incentive-based approach for improving data reproducibility. *Science Translational Medicine* 8, 336 (2016), 336ed5–336ed5. https://doi.org/10.1126/scitranslmed.aaf5003 arXiv:http://stm.sciencemag.org/content/8/336/336ed5.full.pdf

[48] Anisa Rowhani-Farid, Michelle Allen, and Adrian G. Barnett. 2017. What incentives increase data sharing in health and medical research? A systematic review. *Research Integrity and Peer Review* 2, 1 (2017), 4. https://doi.org/10.1186/s41073-017-0028-9

[49] Adam Rule, Aurélien Tabard, and James D Hollan. 2018. Exploration and Explanation in Computational Notebooks. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 32.

[50] Jonathan F Russell. 2013. If a job is worth doing, it is worth doing twice: researchers and funding agencies need to put a premium on ensuring that results are reproducible. *Nature* 496, 7443 (2013), 7–8.

[51] Jonathan R. L. Sears. 2011. Data Sharing Effect on Article Citation Rate in Paleoceanography. *AGU Fall Meeting Abstracts* (Dec. 2011).

[52] N Sadat Shami, Michael Muller, and David Millen. 2011. Browse and discover: social file sharing in the enterprise. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*. ACM, 295–304.

[53] N Sadat Shami, Y Connie Yuan, Dan Cosley, Ling Xia, and Geri Gay. 2007. That's what friends are for: facilitating 'who knows what' across group boundaries. *Proceedings of the 2007 international ACM conference on Supporting group work* (2007), 379–382. https://doi.org/10.1145/1316624.1316681

[54] Victoria Stodden and Sheila Miguez. 2014. Best Practices for Computational Science: Software Infrastructure and Environments for Reproducible and Extensible Research. *Journal of Open Research Software* 2, 1 (2014), 21. https://doi.org/10.5334/jors.ay

[55] Andrea K. Thomer, Michael B. Twidale, Jinlong Guo, and Matthew J. Yoder. 2016. Co-designing Scientific Software: Hackathons for Participatory Interface Design. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16)*. ACM, New York, NY, USA, 3219–3226. https://doi.org/10.1145/2851581.2892549

[56] Theresa Velden. 2013. Explaining Field Differences in Openness and Sharing in Scientific Communities. *In Proceedings of the ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW'13)* (2013), 445–457. https://doi.org/10.1145/2441776.2441827

[57] Daniel M Wegner. 1987. Transactive memory: A contemporary analysis of the group mind. In *Theories of group behavior*. Springer, 185–208.

[58] Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J.G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris a. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3 (2016), 160018. https://doi.org/10.1038/sdata.2016.18

[59] Max L. L. Wilson, Paul Resnick, David Coyle, and Ed H. Chi. 2013. RepliCHI. *CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13* (2013), 3159. https://doi.org/10.1145/2468356.2479636

[60] Daniel J Worden. 2017. Emerging Technologies for Data Research: Implications for Bias, Curation, and Reproducible Results. In *Human Capital and Assets in the Networked World*. https://doi.org/doi:10.1108/978-1-78714-827-720171003