The Value of Data:
Considering the Context of Production in Data Economies

Janet Vertesi
Society of Fellows
Princeton University
Princeton, NJ 08542
jvertesi@princeton.edu

Paul Dourish
Department of Informatics
University of California, Irvine
Irvine CA 92697-3440
jpd@ics.uci.edu

ABSTRACT

In this paper we argue that how scientific collaborations share data is bound up in the ways in which they produce and acquire that data. We draw on ethnographic work with two robotic space exploration teams to show how each community's norms of "data-sharing" are best understood as arising not from the context of the *use* or *exchange* of data, but from the *context of data production*. Shifting our perspective back to the point of production suggests that digital artifacts are embedded in a broader *data economy*. We present implications for analysis of data in interactional context, and for introducing systems or policies that conflict with the value of data in its context of production.

Author Keywords

Data sharing; scientific collaboration.

ACM Classification Keywords

H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms

Human Factors

INTRODUCTION

Scientific collaborations are an area of recent interest in Computer-Supported Co-operative work. A variety of studies employing methods from ethnography to surveys, and enrolling researchers from experimental psychology to HCI to Science Studies, have tackled the question of how "*collaboratories*" [16] work or do not work, and how to design technical solutions to local problems of coordination [2,4,5,6,8,24,32,36,38]. At some point in their understandings about collaboration and principles for system design process, someone on the study team must inevitably approach the thorny question of data sharing.

During WWII and the subsequent Cold War, being identified as "a collaborator" was cause for criminal charges. But early in the 21st century, "collaboration" has come to acquire a positive, and even imperative, tone. A culture of data management associated with networked file sharing, open source software development communities and "the knowledge economy" now dominates popular, corporate, academic and policy discussions about the maintenance and availability of large datasets [1,14]. How much and how rapidly a scientific project shares its dataset has become a metric of success, with agencies such as the U.S. National Science Foundation announcing that *all* funded projects will be required to indicate how research data will be made public [31]. While different groups have different ideas about the nature of sharing and what can be shared, a dominant assumption is that data-sharing is always and everywhere a good thing. Such an attitude has led to a rush in the development of eScience and open platform initiatives for scientists, the assumption being that their participation in an "open society" [29] should make them early adopters in the world of data sharing.

But much social science research has shown that scientists do not, in principle, share their data with other scientists. Nor do they share their data with other scientists in the same way: rather, data-sharing cultures vary across the scientific disciplines [12,13]. Physicists share their data widely, sporting publications with hundreds of authors, while biologists may be more selective about with whom they share, when and why [23,38]. In describing these data-sharing cultures, researchers have focused on how the data is *used* by the collaboration, and how it is *circulated* within and outside the immediate community. Some have examined how cultures of sharing are bound up in systems of credit associated with publication [6]; others have catalogued and described the wide variety of local norms that dominate collaborative work [8,32,38]. Some have examined how disciplinary norms must guide the construction of discipline-specific databases [4,5,12,24], while others have grappled with designing systems and metadata for heterogeneous datasets that arise from interdisciplinary collaborations [3,9,28,36,42]. In general, however, members of the cyberinfrastructure studies community exhort their colleagues to pay attention to the local variations that support both the *use* and *exchange* of data that can make or break a system in implementation. As such, design strategies for data-sharing systems reflect local

sharing norms, and appear to be as unique as the practices at each site of implementation.

This paper presents a framework for how data acquires and exerts meaning within a community, to impact how we study and design cyberinfrastructure for collaboratories: not despite, but because of this heterogeneity. Our case study is a comparative one, based on ethnographic experience with two spacecraft teams: *Helen* and *Paris*.[1] The two missions have much in common: both are prestigious NASA-led projects to other planets in the solar system, featuring large budgets and international partners. Both enroll team members across the United States and Europe, relying on teleconferencing capabilities and networked software to get the work of their missions done. Both are collaborations in the same field, Planetary Science, and even involve some of the same individual scientists. Both are high-profile missions, with broad public engagement. And both are considered extremely successful, publishing hundreds of science papers, making extraordinary discoveries, collecting precious data from distant worlds. For the sake of this paper, however, we emphasize an aspect they do not have in common: their principles and practices of data sharing. In this respect, *Paris* members share raw data internally across all members of the mission and see their data as an inherently shared resource, while *Helen* members do not automatically share internally and are on the whole more selective about where their data may travel.

In this paper, we claim that *Helen* and *Paris*'s different data sharing values do not arise from some attribute of the data's use or distribution among a wider network of planetary scientists; after all, the two missions produce datasets that service and circulate within the same field. Rather, we argue that different values of data sharing arise from the *context of production*: how that data is crafted and acquired. That is, our framework sees data as embedded in a *data economy*, wherein it must first be *produced* before it can be used, stored, accessed, or shared. We begin by examining the organizational structures and work processes on both missions; we then detail how scientists request and acquire data from each spacecraft; and reveal the injurious effects when one system is enforced upon another. Ultimately, we show that *how data is produced* on each mission has pervasive implications for how it may be shared that must be deeply considered by system designers.

Methods
This paper draws on ethnographic materials and methods. The first author spent two years embedded with the *Paris* mission, and a year and a half embedded with *Helen*. The principal observational components included attendance at a full spectrum of planning-related and science-related meetings; interviews with team members at all levels of the organization; and attendance at relevant scientific conferences. Due to the distribution of these teams, the majority of meetings were observed at one of the contributing sites. Site visits to affiliated institutions were conducted on occasion to view the organization from its contributing nodes. During and following transcription, we deployed an open coding scheme: as issues arose thematically in the field site over time, they were tracked in different moments of the organization's work.

In addition to interviews, observations, and field notes, we also produced and analyzed recordings and transcripts of individuals and groups in the conduct of their everyday meetings and scientific work. Parts of meeting transcripts were subject to scrutiny, wherein we took note of actors' accounts of their work, successful arguments for or against cutting or keeping observations, and exchanges that were seen to either breach or support local order. Combining macro-level community participation with micro-sociological analysis, we were better able to characterize the kinds of order produced and maintained within the organizations at multiple scales [18,26].

In our approach and analysis, we draw heavily on the tradition of Laboratory Studies from the field of Science & Technology Studies [i.e. 23,26,38]. Since the 1970's, anthropologists and sociologists have endeavored to understand how science is practically accomplished as a human, social activity. In doing so, they have confronted the popular myths of the "unity of science" or its "normative structure" [29] to show how scientists enlist a variety of practices, social structures, local norms, and other resources in the process of making knowledge about the natural world. Going behind the scenes in the laboratory, then, is not a question of airing dirty laundry, but is aimed at analyzing the *how* of science in order to better understand its role in society, and to design policies or systems that best fit and support how science is done. Similarly, here we use behind-the-scenes examples from two missions to demonstrate not that one is better than or more effective than the other, but to show how the different practices in operation on each mission impact the way in which mission data circulates among planetary scientists, with implications for data-sharing system design and policy.

WHERE DOES DATA COME FROM?
Always referred to by team members in the plural, "*the data*"[2] that spacecraft return in actors' accounts varies in form depending on the instrument. "*The data*" could refer to an image composed of pixels. Planetary scientists take the numerical values of these pixels as direct measurements of photon characteristics and quantity; their analysis of image data can therefore both describe a planet's optical characteristics and to some extent, its composition, atmosphere, morphology or mineralogy. "*The data*" could

---

[1] The spacecraft teams have been anonymized using NASA convention of naming missions after mythological characters.

[2] There is no example of the word "datum" being used on these missions. Consistent with emic accounts of "data", we use the term as they do, as a collective noun.

also be complex graphs generated by spectrometers peering into infrared or ultraviolet light wavelengths, or it could be stellar occultation timings, or measurements of magnetic fields. Such different kinds of data fuel the scientific papers produced by members of each instrument team, who decide which observations the instruments should take to characterize the planet and follow up on their hypotheses. The scientists who work with these datasets use a variety of software suites with many common formats, interfaces, and tools shared across the community. They may download files from a common repository, maintained by NASA, and process images or other datasets in shared systems. But acquiring data from these spacecraft is a deeply social task, achieved through negotiation among scientists and engineers. It is also achieved in different socio-technical architectures, what we might identify as "the human infrastructure of cyberinfrastructure" [25] that dictates how the work of data acquisition gets done.

The Missions' Sociotechnical Infrastructure

Helen is a large mission, with over 10 instruments dedicated to understanding different attributes of its planet under study. These include remote sensing instruments such as a camera, infrared and ultraviolet spectrometer; and physics instruments for measuring planetary fields and particles. Each of these instruments has its own Principal Investigator (PI), all presided over by a Project Scientist. A Project Manager oversees the entire operation. With several hundred participants across the United States and Europe, the organization sports a broad and deep hierarchy: roles are determined within this structure that prescribe how the different parts and people relate to the mission as a whole, and what work is done by whom.

Additionally, *Helen* is a matrix organization. Scientists belong primarily to an Instrument Team, and secondarily to a Working Group related to their area of interest, such as the planet's moons or its atmosphere. Target Teams in each Working Group are responsible for planning observations corresponding to different periods of *Helen*'s workdays specific to that area of interest. Scientists may only request observations from the instrument whose team they belong to. The data that *Helen* acquires is first subject to a "*validation*" period of 6 months or more, and is then released to the NASA database, the Planetary Data System (PDS); all raw images from the camera system are released on the Internet to the public as soon as they are downlinked from the spacecraft. By and large, instrument teams do not share pre-validated data with other teams before its release to PDS, and members of other instrument teams are restricted from publishing about it until that time (an important exception is discussed below).

In terms of structure, *Helen* is fiercely *independent* in orientation. Instrument teams each proposed separate lists of scientific investigations to NASA that they intend to accomplish during the mission period. In *Helen's* Working Groups, a scientist representing each Instrument Team sits on a Targeting Team meeting that meets bi-weekly by teleconference to negotiate with their teammates for the time and bytes to conduct these separate investigations. Instrument Teams, Working Groups, and Targeting Teams each operate independently, often with different behavioral codes or rules and different software suites with varying interfaces to other missions. They exhibit fierce subgroup loyalty and maintain strong subgroup ties. These and other operational procedures will be described further below.

The smaller mission, *Paris*, operates under a single PI. The spacecraft has less than 10 instruments dedicated to different attributes of its planet under study, including an infrared instrument and color camera, a variety of spectrometers, and geologically-oriented analysis tools. With fewer than 200 participants across the United States, Canada and Europe, its structure is that of a flattened hierarchy, with a Principal Investigator and a NASA Project Manager at the top, and all scientists grouped together in what is construed as a single Science Team. Scientists are appointed to manage each instrument, but their roles are oriented towards ensuring the instrument's construction, calibration, health, and commanding, rather than registering authority over other members' requests for observations. Scientists also congregate in Working Groups organized around different areas of interest, such as atmospheres or geochemistry. Any scientist on the team may request observations from any instrument at any time. Data is released on an instrument-by-instrument basis to the PDS after a 3-month validation period, and all raw images are released on the Internet to the public as soon as they are downlinked. All instrumental data is shared internally among all scientists on the team prior to this release.

Comparing the *Paris* mission to *Helen*, the structure is more *interdependent*. The NASA-funded scientific investigations enroll multiple instruments working together to understand the planet. Its Working Groups are responsible for coming up with questions that involve a "*campaign*" of observations across multiple instruments to resolve. As will be discussed below, allocating spacecraft time is accomplished in a daily teleconferenced meeting in which instrument operators, scientists representing each Working Group, and those engineers at the NASA center who command *Paris*, come together to discuss the following day's activities. Throughout, the dominant environment is one of continual, unilateral agreement.

In choosing the terms *interdependent* and *independent orientation*, we purposefully avoid value judgment.[3] Neither system is better or worse than the other, and neither

---

[3] We avoid Hofstede's categories of "collective" and "independent" [21] both because of implied preference, and because we do not believe that these are essentialized notions of cultural difference; rather, as will become clear when we discuss the case studies in turn, such "values" are only visible as they are generatively constructed through practice [22].

can be said to collaborate more or less than the other. The two orientations produce different yet equally complex operational environments, and are subject to different but equally important pressures in the course of their work. Each system arises from and is implicated in the different operational constraints of the spacecraft and their mission type. But in their processes of data collection specific to each mission environment, we see how the missions' work practices constitute two different cultures of data production, and thus exchange. As we shall see, local meanings and practices of "collaboration" and "sharing" vary in these two cultural contexts, exposing how our very idea of "sharing" and "collaboration" must change from context to context to remain sensitive to local norms.

Allocating bits

Spacecraft time, bits, and power are tightly constrained on all planetary science missions, making them resources that are bound to produce task conflict [35]. Each mission has evolved different techniques and technologies to manage and resolve such conflicts. On both *Paris* and *Helen*, scientists participate in rigorous planning groups in which they must propose and advocate for an observation and see it through to selection and coding for implementation. In both cases, however, the independent and interdependent orientations produce and reflect very different operating procedures and decision-making cultures.

Planning for Paris: Interdependence and Communalism

As outlined above, *Paris* brings together representatives from each Working Group, operators of each instrument, and the engineers who command the spacecraft, to meet together daily to decide on the spacecraft's next step. The kick-off meeting for each sequence of spacecraft commands is a highly ritualized practice with a clear order of events and expectations of attendees. The goal of the meeting is to come to unilateral agreement on what the robot should do the next day. Should a team-member disagree, the process could be derailed until they are on-board again.

This consensus expectation pervades the team, not only establishing the ground rules of interaction, but also crafting an initial mindset or orientation towards the day's activities. This is consistent with experimental findings that consensus-orientated groups display greater "cognitive consensus" at the outset that structures their subsequent interactions [30]. This begins at the outset of the meeting as a mission member reviews where the spacecraft is and reminds the team what they intended to do at that location. It persists in the use of the word *"we"* to elide both the spacecraft and the human team, and enrolls members of that team into a single dominant identity. When observations are "*cut*" to satisfy the restricted number of bits available that day, the team considers this a success, "*declaring victory,*" as they have collectively made difficult decisions about their robot's activities. Indeed, the work of the mission meeting is one of ensuring constant assent and engagement.

Throughout *Paris*'s planning meeting, as observations are requested by scientists, they are added to the software by a meeting participant who acts as the "*Keeper Of the Plan.*" The home-grown, networked software shows a running tally of the number of bytes and watts used, such that all members monitor commands. When the plan is approved, individuals go on to code the instructions for their instrument. But beforehand, due to the visibility of all observations to all members, the process enrolls collective oversight and self-discipline to "*mind the bit bucket*" as members slot observations into sequence.

Thus participating in the interdependent *Paris* mission requires setting aside individual goals to join the group. Roles with strong responsibilities are allocated among the science and engineering teams, but these roles rotate: thus team members may make tough calls without taking on a personal loss of face among the community. All members are considered equal, restricting individuals from exerting authority over their colleagues. Shared software, shared data, shared outlook: the interdependent outlook pervades every aspect of mission operations and engagement.

Planning for Helen: Independence and Fairness

Unlike the unifying, collectivist space of *Paris*, *Helen* members must "*wear multiple hats*", moving between their memberships on an Instrument Team, a Working Group, and a Target Team, and representing the needs of each to the other. Scientists belong first and foremost to Instrument Teams wherein they are charged by their PI or leader to complete an observational or analytical task often outlined in the instrument's proposal; they discuss their evolving interests in one or another facet of the dataset with their colleagues on that instrument team and may pursue investigations of their own. Instrument PIs work hard to protect their team members' scientific questions from outsiders who may become interested in the same facet and "*scoop*" their teammates' findings.

Scientists who are interested in particular observations that have been approved by their Instrument Team must look over the spacecraft's planned trajectory and input a proposal for the command into a Mission-wide database. This database produces outputs for each segmented period showing when scientists on different instruments (or sometimes different scientists using the same instrument) input requests that overlap. These become the basis for negotiation, during the bi-weekly Target Team meetings, as to which observations will go forward and which will be cut. All team members can see which observations are in conflict, although not necessarily who "*owns*" (inputted) each observation. Early in the mission the software was modified to allow aspects of the observation to be hidden from members of different instrument teams, to protect individuals' experiments; this feature is now rarely used.

The Target Teams have worked together for many years, and have each developed unique methods of resolving observational conflicts. Some meet far in advance and

allocate each segment to each instrument in turn, such that the main arguments as to who may conduct what science observations are resolved before proceeding to individual planning meetings. Others request that Instrument Team representatives decide which observations to put forward from their team; still others request that all scientists individually enter all their requested observations into the database for representatives to negotiate during the Target Team meeting. The techniques for decision-making vary: as one engineer who had recently moved from one Target Team to another explained: *"When I tried Target Team A's 'percentage cuts across the board' [method of byte-trimming] in Target Team B, people complained... In Target Team A, I once tried the volunteered cuts, Team B's method, and ... [the chair] noticed and said "Did we change our procedures?! ... I didn't [know] that the conventions in each group were so ingrained."*

In addition, Target Team representatives must creatively balance fluency with local procedures and the pressure to secure spacecraft time for their and their Instrument Team colleagues' observations on the one hand, with the technical constraints of the spacecraft on the other. Not all observations are feasible. Due to the spacecraft's construction, *Helen* can generally only point one instrument at a time at a given target, although some specific combinations of instruments are possible. Further, the spacecraft must turn between different observations to adjust its alignment for each instrument, and is often commanded to roll or look away from the sun at various points in its path; this requires precious fuel and places strain on the turning mechanisms. As these resources must be preserved, Target Teams must come up with a plan that is feasible and safe for execution, and that balances Instrument Teams' science observation requests.

Team Norms in Planning

Whereas planning on the interdependent *Paris* mission requires maintaining the team's collective orientation, planning on the independent *Helen* requires respecting the autonomy of subgoups. The different orientation produces different affective experiences and rewards different interactional strategies on both missions. *"It always helps to sound friendly,"* explained one member of *Paris* as he recounted what it took to get an observational request approved by the group; and another informed us that he consciously made an effort to make himself sound more open and lively on the telecon line to remind his colleagues that he was being co-operative. Even if your observation is cut, or if you do not agree with a particular decision, a tone of acceptance and agreement will eventually be assumed. This smoothes over conflicts, but can also make it difficult for individuals who disagree to get their point across. To discourage "silencing" [34], which can readily occur in teams that enforce such conformity among members, the team has taken a strategy of active listening to heart: *"[Your colleague] could be wrong 99% of the time, but if it's that one percent, you'd better be listening,"* one recounted. But

this strategy often means that the team attempts to appease as many of its members as possible by frequently choosing to honor dissident voices, even if they are in the minority. Common sense on *Paris* dictates that it is better to maintain an environment of collective agreement than to turn any individual down, as this might encourage factionalization.

As they wear multiple hats, negotiate different subcultures and resolve difficult choices between independent instrument teams, *Helen* members display a direct approach to managing conflict. Members never hesitate to state their needs, take a declarative tone of voice, and press their case. This is essential to conducting their science, as members of other subgroups may not understand the rationales for local decisions or needs. Unlike a *Paris* team member's ability to appeal to a collective, or to role-play and downplay their individual role in a conflict, the independent approach on *Helen* may require team members to put their face on the line in high-intensity interactions [19]. Despite the use of jokes to assuage these interactions, some participants do *"develop a reputation"* among their peers, leading *Helenites* to note, *"We've got a lot of personalities on this mission."*

A final point about data production on both teams: on *Paris*, scientists are expected to give up individual requests for requests that satisfy the goals of the team. When those needs have been met, even if that means sacrificing observations, team members call consensus and declare themselves to be *"happy."* If individuals are begrudging about such choices they may rarely show it to their teammates. In their philosophy of science, *Paris* team members frequently express that *"what the team decides [together] is what's best for the spacecraft,"* and/or *"best for the science."* This point of view clashes with that on *Helen*, where the bold, direct method of negotiation is seen as better for mission science. As one *Helenite* said: *"I don't think the process of science is well served if a bunch of people get into a room and say, let's agree not to disagree."*

To resolve conflicts arising from incommensurable observational needs and constrained spacecraft pointing opportunities, then, the *Helen* team has developed an overriding concern for *"fairness."* That is, no team should be seen to be treated unequally; every team must have its fair share of the spacecraft's time to complete their science goals and acquire their data. As one *Helen* team member reminded her colleagues during an argument over two competing observations, *"Our Instrument is not the only Instrument and our Target Team is not the only Target Team."* This allows the group to shy away from judging the value of each other's science, with which they may be unfamiliar. Consistent with organizational justice research [15], exactly how fairness is achieved differs from subgroup to subgroup, leading to subculture clashes when two groups don't see eye to eye on a conflict or its method of resolution. But despite (or perhaps because of) this interactional intensity, *Helenites* are quick to engage socially, are passionate about their mission, and relish each other's company, calling their teammates *"family."*

DATA CULTURES

The purpose of the above section is emphatically not to extol the virtues of one or another mission. Instead, if we hold both symmetrically, we must note that scientific production, team membership and collaboration all look radically different on each team. This fact has been discussed in the literature before with respect to different fields and their contrasting methods of data sharing [8,23,36,38]. However, *Helen* and *Paris* are collaborations within the *same* field. While the patterns and expectations of data use and circulation are the same, we will discuss how their data sharing practices are in fact very different. They are more consistent with the aspects of data *production* described above, than they are with differing disciplinary cultures of data use or publication rights.

Until recently, data-sharing practice in Planetary Science followed the model adopted by *Helen*. Each instrument is separately commissioned by NASA in an initial competition: PIs have a responsibility to manage their instruments "*from the cradle to the grave*," including planning for and acquiring observational data. For the most part, scientists on the mission belong to one of these Instrument Teams and only have access to data provided by that team.[4] Additionally, each instrument team is responsible for releasing their calibrated data, raw data and calibration algorithms to the central NASA database for all scientists, within and outside the mission, to use. Many instrument teams employ strict protocols, known as "*Rules of the Road*," with respect to the management of publications, credit, and sharing with outsiders. With the strong exception of the physics instruments, discussed below, the majority of *Helen* team members' publications are single instrument-focused investigations of one aspect of the planetary system. Deep datasets develop for each instrument, and each team is at leisure to develop instrument-specific suites of observations aimed at the pursuit of a scientific problem.

On *Paris*, the interlocking instrument suite reflects and promotes the interdependent attitude. Observations on *Paris* are crafted to be multi-instrumental, to achieve team-wide science goals, and are the communal property of all scientists on the team. One scientist stated what, after so many years on the mission, was simply obvious to her, "*You shouldn't limit yourself to one instrument, it's the most foolish thing you can do!*" The data files and display software are written to be combinable. Radical interoperability is practiced: between instrumental datasets on the one hand, and between scientists in their collective team on the other. Scientists are required to present results-

in-progress to their colleagues and circulate papers in early stages of their development to open authorship to any team member who wants to participate. Interoperability has gone so far on *Paris* that second-order visualizations have developed that show data from multiple instruments as already-combined: e.g. mineral weights, percentages, and composition, factors detected by three different instruments, can be displayed in a single graphic form.

On the more independently-oriented *Helen*, "*co-ordinated observations*" may present "*a major political issue.*" Such observations require considerable work as they breach Instrument Teams' local norms, and strain co-ordination points and loyalties between Instrument Team members at all levels of observation development. For example, a multi-instrument observation requires co-ordination between scientists on different teams with different subcultures, norms, and ways of resolving challenges; between instrument operators at different remote institutions whose software suites may be incompatible; between instruments bolted to the *Helen* spacecraft that may not physically be pointed at the same target at the same time; and between members of a Target Team or an Instrument Team who have different ways of assessing whether or not a set of allocated observations is "*fair*". Importantly, however, *Helen* also achieves certain science goals that are difficult to achieve on *Paris* because they violate *Paris*' operating procedures: for example, building "*deep*" datasets such as catalogs of single-instrument observations of the same features during different seasons, or limiting publication author lists to showcase a single team member's or Instrument Team's dedicated work.

When Cultures Collide

If the story ended here, we might see simply two different collaboratories working with data in different ways. But while *Paris* and *Helen* operate in the vacuum of space, they do not operate in a vacuum on Earth. The expectations of data sharing have changed since *Helen* was designed. This has resulted in the imposition of *Paris*-like rules for data sharing upon *Helen*, producing culture clashes and breaches to existing order on *Helen*. The confusion, frustration and sometimes resentment that results from such attempts reveals important considerations for collaboratory policy-making and system design.

The first such breach occurred when, before *Paris*' launch but after *Helen*'s, *Paris* team leaders went to NASA headquarters with a bold proposal: all their raw image data would stream directly to the Internet, so that members of the public could follow along with their mission. *Paris* leaders recount the astonished looks on the engineers' faces when they demanded assistance in making their data instantly available to the public and to each other. NASA Headquarters, looking for ways to engage the public and demonstrate the value of their missions to Congress, was so delighted that it required other missions, including *Helen*, to follow suit. NASA argued that since the instruments were

---

[4] Two exceptions exist to this rule. First, a handful of "interdisciplinary" scientists were selected such that they may request data from more than one instrument, but they must enter into agreements with each instrument's PI and follow the rules of each Instrument Team in order to do so. Second, the physics instruments behave more like *Paris*, as will be discussed below.

taxpayer-funded, it was only right that taxpayers could see for themselves the fruits of their dollars.

The *Helen* team complied, but the levied requirement proved internally divisive. After all, unlike *Paris*, no other dataset is shared among team members before the end of the validation period. Because missions like *Helen* require an independent orientation, other members of the mission may represent competition for spacecraft time, observations and discoveries. The negotiations in which *Helenites* have to engage at all levels of the data-production process register a personal investment that is considered fairly repaid with a validation period to protect publication rights. Mandated, gratuitous, or unapproved sharing of data inside or outside the mission before the end of the validation period therefore breaks the team's Rules of the Road and is seen as "*unfair,*" compromising members' ability to do the jobs they were selected and funded to do. While they work to make their calibrated data understandable to expert colleagues, *Helenites* continue to be cautious of attempts to make their data "*too easy*" for others to use, even scientists on their own mission, without the deep, trusted knowledge of an instrument that can come with participation on a team.

The data culture clash evident above is not always between two missions: it may exist on a single mission alone. For example, the *Helen* Physics Working Group subgoup displays a subculture and orientation much like that of the *Paris* team. Because questions about a plasma field or an ionized discharge from a planet require multiple sensors to answer, the *Helen* physics instrument suite works in a complementary fashion. A database allows members of other Instrument Teams to access lower-resolution versions of each others' data; the group has a communal Rules of the Road document to govern data sharing and publication; and publications include authors across multiple instruments. This observed interdependent approach is consistent with studies of other particle physics collaborations [23,38].

In 2009, the members of the Physics Working Group on *Helen* invited members of the other Instrument Teams to join them in a new Science Working Group devoted to the exploration of a newly-discovered phenomenon. As the Group's leader explained, "*multi-instrument observations are important to achieve the next level of science*." Thus, at the initial meeting, the Chair opened discussion with a proposal: "*I would like to put a group together and to basically get together and come up with the information, observations and access to data that are needed to do the best science for [the Phenomenon]."*

This proposal was met with confusion. Members of newly-added Instrument Teams were unclear as to what else was requested of them, and what they could expect in return. In case they were being accused of not sharing, they reminded their colleagues of their openness to collaboration as evidenced by their presence at the meeting, and their willingness to accommodate other instruments' requests in planning their observations, should they ask. Members of

the Physics Working Group insisted that they already had a strong record of collaboration within their Group, insisting, "*you're fixing things that aren't broken!*" As the discussion wore on, other members expressed concern about goals and process: "*I'm still not clear what we're trying to do and how we're trying to do it, and how won't make sense unless I understand what.*" The Group Leader wanted to initiate a "*new process*" that would produce data that the group considered to be shared, but the assembled group quickly pointed out that his "*new system*" violated too many of their existing rules of engagement, from incommensurate subgroup Rules of the Road to processes of data acquisition and management. When a senior scientist in the room finally started singing "*Kumbaya*", the group dissolved into good-hearted laughter at their predicament.

Following this initial meeting, the group met in person twice more. While the "*new processes*" were never explicitly worked out and no top-down notion of required sharing was established, a few bottom-up initiatives that the group approved as indeed "*collaborative*" started to take shape. For example, two scientists belonging to different observational instruments began to work together (for the first time in six years of operations) to figure out how to co-ordinate observations between the ultraviolet and infrared spectrometers. This was complicated as each instrument not only pointed in different directions, but required the spacecraft to behave differently when taking each observation. Still, the two were prepared to compromise, explaining: "*what's ideal for [them] turned out to be not very good for [us], but what's fairly good for [them] turned out to be fairly good for [us].*" Thus the two teams met in the middle, both accepting a "fair" downgrade to their data in the interest of an alternative approach to data collection.

This example neatly reveals how different ideas of what "*collaboration*" and "*sharing*" mean in the first place were already in operation in each of the subteams. *Helen*'s independent orientation had produced different cultures of data acquisition and management, with different notions of what it means to "*share*", with whom it is possible to "*collaborate,*" and when such collaboration should take place. When the call to "*collaborate*" and/or "*share*" was heard, each group asserted that they were already doing it, and tried hard to understand if something different was required of them. It took three or four meetings before a shared notion of "*sharing*" took hold and new practices of data acquisition started to emerge.

DISCUSSION
Without knowledge of how the two missions work, the data sharing regimes and culture clashes described above are impenetrable, perhaps ascribed to generosity or selfishness. But knowledge of *how the two missions produce data* in the first place makes sense of the two missions' opposing data-sharing cultures. Both are planetary science ventures; both have scientists who prefer to *use* the data they collect in very similar ways; both are also considered highly

successful missions. The key difference that affects sharing regimes is the organizational culture that dominates data collection. The CSCW community is already familiar with the idea that organizational culture and structure determines how or whether a new technology may be adopted [20,33]: in this case, we see how such organizational considerations affect *how the collaboratory collects data*, influencing how scientists share that data downstream. It is not the structure or culture of data *sharing*, then, but the structure and culture of data *production* – how the collaboratory collects data – that determines the parameters for sharing, and must inform our design interventions.

This focus on the organizational aspects of data production explains why the transportation of *Paris'* models of data-sharing to *Helen* (or even from one subgroup of *Helen* to another) produces complications: the datasets have fundamentally different values. Data on *Paris* is not a shared resource simply because it is data: it is a shared resource because *it is crafted that way from the outset*. It is circulated within a community whose every practice of spacecraft management is simultaneously geared towards the maintenance of an open (albeit bounded and tightly regulated) society. Observations on *Paris* may be suggested by an individual, but they reflect the decision-making process of an interdependent collective. Further, instrument data on *Paris* is already inherently combinable. The ability for all scientists to use all instruments has resulted in the asking of science questions that *demand* the use of multiple instruments in their answering. The culture of interdependence and collectivism does not allow for any one scientist to call any observation, "*my data.*" Thus sharing *Paris* data among team members or even distributing it beyond the team is not a cause for team members' concern: the data is already and always understood internally to be a shared resource.

Data on *Helen*, however, has a different value. It is the product of individuals creatively resolving inter-team conflict while respecting subteams' autonomy and authority over their own production. Observations are hard-won by representatives in Target Teams, often in direct conflict with their colleagues, who had to decide how to fairly allocate spacecraft time to achieve multiple competing and irreconcilable goals. Such high-intensity interactions are balanced with subgroup loyalty and strict adherence to and respect of local cultures of conflict-resolution. But when conflict becomes severe, personal reputations are at risk within the wider community. Thus the data that returns from *Helen* is *not* a free, shared resource, but an expensive and hard-won one, representing the work of independent, autonomous teams. It is very much "*mine*" or "*yours.*"

It is impossible in a ten-page paper to thoroughly discuss the relationships, processes, individuals and stories that give membership on *Paris* and *Helen* meaning. As one team member reminded us, "*Missions, like people, are complex entities with many subtleties, strengths and weaknesses.*" However, we emphasize again that our telling of this small part of the story is not meant to establish praiseworthy or problematic models for collaborative work; nor is it meant to establish a taxonomy of collaboratories with greater or lesser propensity for sharing [as in, e.g., 8]. Instead, by framing the value of data as established through the context of its production, we present a new understanding of data-sharing as part of a *data economy*, with implications for both cyberinfrastructural design and collaboratory policy.

Confronting the Data Economy

We propose that data-sharing is only one set of practices in a larger *data economy* that encompasses production, use, and circulation. While the latter has traditionally been of interest to those interested in eScience and collaboratories, who characterize a *data ecology* in how data moves among and between users [39], the above cases make clear that the former is also of tremendous importance to the community that produces the data for others to use. Indeed, "ecologies" of enmeshed sociotechnical infrastructures for data-sharing are part of this wider economy of data, but attention to other aspects of this data economy – production alongside use and exchange – is critical for designing data infrastructures that reflect and secure data sharing practices specific to each scientific collaboration. This formulation of value as determined by context of production is useful to the collaboratories analyst in three ways.

First, this framework encourages that the CSCW community come to see data in **interactional context**. That is, *work with data enacts social relationships:* data is one of the resources or interactional elements with which social relationships are performed [7,19]. In the context of data production, we witness different interactional regimes at play in these two missions that orient and establish membership. Willingness to share and complicity in collective data management is required of *Paris* scientists. In-team sharing of results and collective interpretation of their spacecraft's data is what makes them part of their team. To keep one or another piece of data or a discovery to oneself would be to act outside the team's repertoire of interactions, even taboo. Performing membership on *Helen*, instead, includes entering into frank negotiations with other team members, loyalty to one's own Instrument Team or Working Group, and respect for each subgroups' autonomy, independence, and local norms.

While the *Paris* interactional regime and systems of data exchange mutually support their collaborative culture, when moved to the other side of the solar system they deeply disrupt the practices and interactions that give *Helen* team data meaning to its members, and establish membership. This clash of data infrastructures produces "torque" among its membership [10], as the technology and policy demands can restrict members from enacting their very membership within the group. This perspective explains how *Helen* scientists' resistance to the importation of a *Paris* model of sharing is not due to cultural conservatism, but rather to stark incompatibilities with their mission's existing human

and technical infrastructure and socio-material practices [for parallels in other domains, see 41].

Attention to how interactions and social relationships in the context of production produce value, recalls the classic formulation of value established by Marx [27]. According to Marx, the labor and social relations required to produce an object determines its initial value, but this value is obscured when the object moves into the market where it is "commoditized", acquiring use-value (determined by demand) and exchange value (use-value in monetary terms). It is not the goal of this paper to produce a deep Marxist analysis, nor to inspire recourse to other theories in economics. However, thinking analogically, the data economy framework offers a new way to approach problems in the imposition of those data-sharing systems and regimes that do not respect the context of production.

Second, then, this framework reveals that the "*data*" in "*data-sharing*" has undergone a troublesome process of **commodification**: a process that obscures the interactional and social regimes that produced it and give it meaning. While we usually associate "commodification" with cost inflation, the economy into which planetary (and other kinds of) data has entered has deflated its value to that of "free." Mistaking the exchange value of *Paris* data (which is consistent with its production value) for the value of *all* planetary science data, NASA determined a value for *Helen* data that is both distinct from and highly disruptive to its production value, with associated interactional norms, relationships and procedures. But it is not only NASA that enforces a disruptive value for planetary data. System designers and architects of "the knowledge economy," informed by their own historically-situated and discipline-specific values, may still display an overriding belief that "*information wants to be free*." [1,14] The imposition of the value of "*free*" – or "*open*" or "*shared*" – upon a dataset that does not have this value in production, opens those who attribute a commodified value to a product to hostility from those most closely familiar with the data from the site of its context of production. Any "*problem*" with data-sharing on collaborations like *Helen*, then, occurs when external agents exert pressure on the team to adopt a commoditized value for their data inconsistent with its context of production. Such a move may disrupt interactional norms and acquisition procedures, and deprive community members means of asserting their membership.

Thirdly, attention to context of production in data economies reveals how data **fetishization** may easily plague database design [37]: that is, when an object's value is over-determined by its context of exchange, this single object may stand in for and obscure the social relationships that initially gave it meaning in its context of production, acquiring value or exerting agency by other means. In the case of *Helen* and *Paris*, the rich social practices and relationships and the material conditions of data-production that give that data meaning for its community of users, are lost in the circulation of the planetary science dataset. In the process, the dataset comes to represent the entire *Helen* or *Paris* mission, even speaking directly for the planets under study, bypassing its deep sociotechnical origins. In doing so, it gains in scientific authority and credibility. But what the dataset gains in authority is traded at the obscuration of the socio-material context of its production. As analysts and system designers, we ignore this context at our peril, as violation of its rules of engagement may inspire strong resistance among the communities we aim to serve.

CONCLUSIONS

The shift to a *data economy* framework exposes a new and rich avenue for research in the development of collaborative systems. After all, much work in computer-supported cooperative work has fruitfully looked to what scientists need to *do* with data – search for it, combine it, share it with some people but not with others, archive it, manage multiple ontologies for it, etc. – in order to design computational solutions. Other fields of system design and policy-making still display a fascination with data as a commoditized and even fetishized object: something that should be always, and everywhere, available, free, open, accessible, translatable, and searchable. Such policy-makers and designers have produced systems or pronounced degrees of access consistent with this perspective. But as this example of *Paris* and *Helen* makes clear, it is only through observing *the context of production* that we can understand a critical aspect of datasets: the value they acquire from the community that produced (and uses) them.

We must also realize that "data sharing" can only sensibly be described as "sharing" from within the values framework of a very particular system of exchange. Articulating different *economies of data*, involving production alongside use and exchange, may help us to better understand how data acquires meaning in scientific communities and beyond. This would assist in clarifying the "value" of data as distinct from "values" in design [17]: attributable not primarily to cultural values writ large but rather to local organizational cultures. It would also clarify how a request that members of an organization has restricted access to others' data is not actually a question of "privacy", but rather arises from existing organizational "accountabilities" [11,40]: social relations that data exchange enacts. It also provides guidelines for nascent communities to design data production practices that are consistent with the desired circulation outcomes. The technological infrastructures that we introduce to each new information-economy context must respect and enhance – or at least, not directly challenge – the processes by which the data they handle gains currency and value: including those specific to the context of data production.

ACKNOWLEDGMENTS

REFERENCES

1. Anderson, C. *Free: the Future of a Radical Price*. Hyperion, New York, NY, 2009.

2. Aragon, C., Poon, S. and Silva, C. The Changing Face of Digital Science: New Practices in Scientific Collaborations. *Ext. Abstracts CHI 09*, ACM Press (2009), 4819-4822.

3. Baker, K. S., Bowker, G. C., and Karasti, H. Designing an infrastructure for heterogeneity in ecosystem data, collaborators and organizations. *Proc. Digital Government Research* (2002), 1-4.

4. Bietz, M. & Lee, C.P. Collaboration in Metagenomics: Sequence Databases and the Organization of Scientific Work. *Proc. ECSCW 2009,* Springer-Verlag (2009) 243-262.

5. Birnholtz, J. and M. Bietz. Data at work: Supporting Sharing in Science and Engineering. *Proc. Group '03*, ACM Press (2003), 339-348.

6. Birnholtz, J. What does it mean to be an author? The intersection of credit, contribution, and collaboration in science. *J. Am. Soc. Inf. Sci. Technol.* 57,13(2006), 1758-1770.

7. Blumer, H. *Symbolic Interactionsim: Perspective and Method*. University of California Press, Berkeley, CA, 1969.

8. Bos, N., et al. (2007) From shared databases to communities of practice: a taxonomy of collaboratories. Journal of Computer Mediated Communication, 12(2), article 16

9. Bowker, G. Biodiversity Datadiversity. *Social Studies of Science*, 30, 5 (2000), 643-683.

10. Bowker, G. and Star, S.L. *Sorting Things Out*. MIT Press, Cambridge, MA, 1999.

11. Button, G. and Sharrock, W. The Organizational Accountability of Technical Work. *Social Studies of Science* 28 (1998): 73-102.

12. Carlson, S. and Anderson, B. What are data? The many kinds of data and their implications for data re-use. *Journal of Communication*, 12, 2 (2007), 15.

13. Ceci, S. Scientists' attitudes toward data sharing. *Science, Technology & Human Values*, 13 (1988), 45-52.

14. Christen, K. Access and Accountability: The Ecology of Information Sharing in the Digital Age. *Anthropology News*, "Visual Ethics." (April 2009), 4-5.

15. Colquitt, J. et al. Justice at the Millenium: A Meta-Analytic Review of 25 Years of Organizational Justice Research. *Journal of Applied Psychology*, 86, 3 (2001), 425-445.

16. Finholt, T. A. Collaboratories. *Annual Review of Information Science and Technology* 36 (2002), 73-107.

17. Friedman, B. Value-sensitive design. *interactions* 3, 6 (Dec. 1996), 16-23.

18. Garfinkel, H. *Studies in Ethnomethodology*. Blackwell, 1967.

19. Goffman, E. *Interaction Ritual*. Aldine, Chicago IL, 1967.

20. Grudin, J. Why CSCW Applications Fail: Problems in the Design and Evaluation of Organizational Interfaces. *Proc. CSCW '88,* ACM Press (1988), 85-93.

21. Hofstede, G. *Culture's Consequences*. Sage, Bev. Hills, 1980.

22. Irani, L., Vertesi, J., Dourish, P., Philip, K., and Grinter, R. E. Postcolonial computing: a lens on design and development. *Proc CHI '10*, ACM Press (2010), 1311-1320.

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

24. Lee, C. P., Bietz, M. J., & Thayer, A. Research-driven stakeholders in cyberinfrastructure use and development. *Proc. 2010 International Symposium on Collaborative Technologies and Systems*. IEEE Press (2010), Washington, D.C., 163-172.

25. Lee, C. P., Dourish, P., and Mark, G. The human infrastructure of cyberinfrastructure. *Proc. CSCW '06*, ACM Press (2006), 483-492.

26. Lynch, M. *Scientific Practice and Ordinary Action*. Cambridge University Press, Cambridge, MA, 1997.

27. Marx, K. *Capital*. Penguin, London, 1867(1976).

28. McDonough, J. Structural Metadata and the Social Limitation of Interoperability. *Balisage: the Markup Conference Proceedings* (2008).

29. Merton, R. The Normative Structure of Science. In N.W. Storer (ed.) *The Sociology of Science: Theoretical and Empirical Investigations*. University of Chicago, 1942(1973).

30. Mohammed, S. and Ringseis, E. Cognitive Diversity and Consensus in Group Decision Making. *Organizational Behavior and Human Decision Making*, 85, 2 (2001), 310-335.

31. National Science Foundation (2010). Press release 10-077.

32. Olson, G., Zimmerman, A. and Bos, N. *Scientific Collaboration on the Internet*. MIT Press, Cambridge, 2008.

33. Orlikowski, W. Learning from Notes: Organiational Issues in Groupware Implementation. *Proc. CSCW '92,* ACM Press (1992), 362-379.

34. Perlow, L. and Repenning, N. The Dynamics of Silencing Conflict. *Res. Org. Beh*. 29 (2009), 195-223.

35. Pondy, L. Organizational Conflict: Concepts and Models. *Admin. Sci. Quart.* 12, 2 (1967), 296-320.

36. Ribes, D. and Bowker, G. Organizing for Multidisciplinary Collaboration: The Case of the Geosciences Network, in Olson, G., Olson, J. and Zimmerman, A. (eds.) *Scientific Collaboration on the Internet*. MIT Press, Cambridge, MA, 2008, 311-330.

37. Shklovski, I., Vertesi, J., Troshynski, E., and Dourish, P. 2009. The commodification of location: dynamics of power in location-based systems. *Proc. Ubicomp '09*, ACM Press (2009), 11-20.

38. Shrum, W., Genuth, J. and Chompalov, I. *Structures of Scientific Collaboration*. MIT Press, Cambridge, MA, 2007.

39. Star, S. L. and Ruhleder, K. Steps towards an ecology of infrastructure. *Info. Sys. Res.* 7 (1996), 111-134.

40. Troshynski, E., Lee, C., and Dourish, P. Accountabilities of presence: reframing location-based systems. *Proc. CHI '08*, ACM Press (2008), 487-496.

41. Verran, H., Christie, M., Anbins-King, B., Van Weeren, T., and Yunupingu, W. Designing Digital Knowledge Management Tools with Aboriginal Australians. *Design Creativity* 18, 3 (2007), 129-142.

42. Zimmerman, A., et al. The promise of data in e-research: many challenges, multiple solutions, diverse outcomes. In N.W. Jankowski (Ed.) *E-Research: Transformation in Scholarly Practice*. Routledge, New York, (2009), 222-239.