Introduction

Making Sense of Data-Centric Socialities

Scientists need data. Data are scientists' records of observation, collection, and experimentation. Data are the materials that scientists use to construct empirical knowledge. But data also need scientists. Whether in the laboratory or the field, scientific equipment and detectors do not work by themselves. Someone builds, maintains, and operates these recording devices. Someone calibrates, analyzes, and interprets their output, and someone archives these records. Yet there is another, deeper sense in which data, if good to use and good to reuse, need people – scientists, technicians, and administrators – as social beings: embodied humans who communicate; share cultural, normative, and ethical notions; monitor other's actions; build trust; hold others to account; and navigate potential challenges to their reputation. Science may be exceptional because of its systematic ways of learning about the world,[1] but scientists and technicians are in many ways not different from ordinary people. Some studies of science with large datasets broadly acknowledge cultural and social influences, but there is much more to learn when we engage a refined view of what people are like and how they act. We can do this by witnessing scientific work ethnographically, paying close attention to interactions between scientists, students, and technicians.

But as ethnographic fieldworkers we cannot study science in general. We can only witness *this* work in *this* discipline, done *here* and *now*. To study data-rich science ethnographically, we need a science and an example of its workings. *How Data Need People* focuses on astronomy, which is arguably not only the oldest science, but has also been at the forefront of new developments in working with large datasets, the open access to data and software, and uses of machine learning.[2] Allow me to declare right away that

[1] That science is characterized by its systematicity is a view articulated by philosopher Paul Hoyningen-Huene (2013).

[2] Astronomical data are almost entirely digital and largely homogenous – recorded from basically the same viewpoint (planet Earth and nearby spacecraft) with a limited set of detector technologies – are

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 12 Nov 2026 at 16:31:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.002

many of the practices and resources that we find in data-rich astronomy are similar to those in other sciences, making many of this study's findings transferable to them. So let us begin by considering one astronomical dataset and two scenes of its production and publication in some detail. In the rest of this introduction, I will keep returning to this case as I specify my anthropological approach, review my argument, and introduce my study.

I.1 Scientific Data Are Epistemic and Social Stuff

Figure I.1 shows a photograph of a small patch of the night sky, made using an infrared camera attached to the 3.5-meter telescope of Calar Alto Observatory in the mountains of Andalusia, Spain. Taken in what astronomers call the H band, this is a recording of near-infrared light invisible to human eyes.[3] It is a composite image, a so-called mosaic, the result of adding 580 exposures totaling 10 hours of "integration time." Collected in several observing campaigns over 3½ years, these exposures were recorded using a detector with 2048 × 2048 pixels, cooled with liquid nitrogen to –120 degrees Celsius. The image is a photographic negative. Small black and grey dots, about 32.000 altogether, are traces of light attributed to distant galaxies, some estimated to be 9 billion light years away, two thirds of the way to the end of the visible universe.

Making this mosaic was a distributed and collaborative effort. The observatory's staff astronomers recorded the exposures (at times in the company of a visiting graduate student) and sent them to researchers in Heidelberg, Germany, largely on mobile hard drives.[4] There, members of a research team processed, calibrated, and added them. This mosaic was prepared mostly by Nadine,[5] a PhD student. Senior scientists instructed her to assemble a consistent set of exposures from various observing campaigns, to detect objects and to combine these data with those from an earlier project to measure the distances and physical properties of galaxies in this field. Because of the high data rate of many short exposures taken with a large detector array, Nadine processed these exposures semi-automatically using a so-called pipeline. The target of these observations was A2713, a

rarely monetized, lack economic uses, and come with few privacy concerns. This makes the social study of their use in certain ways less complex and may help to foreground elementary issues and practices.

[3] The H band is the near-infrared wavelength range of 1.5 to 1.8 micrometers.

[4] At times, it seems, data need people for transportation and delivery.

[5] When appearing by themselves, all first names in this book are pseudonyms. Combined first and last names are actual names.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 12 Nov 2026 at 16:31:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.002

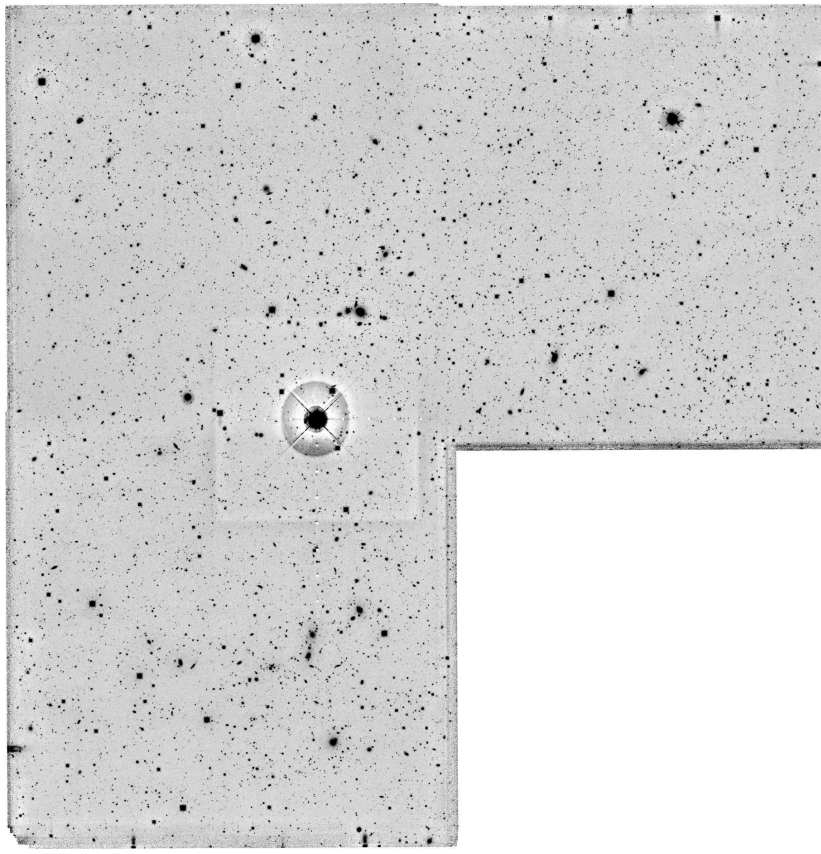


Figure I.1 Mosaic of photographic exposures of the A2713 field.
(Reproduced with permission).
Note: The online version shows the colors of the original figure.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 12 Nov 2026 at 16:31:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.002

supercluster (a "cluster of clusters") of galaxies, as well as galaxies in the universe behind it.[6]

Nadine belongs to MAMBO (pseudo-acronym), a research group that investigates deep fields: small parts of the sky selected for offering views of the distant universe which are little obstructed by the stars, dust, and gas of our Milky Way galaxy.[7] Taking particularly long ("deep") exposures of these fields, these astronomers aim to detect and characterize faint and distant objects. Because of the finite speed of light, we see distant objects as they appeared in the remote past. Equipped with this sort of time machine, the group pursues what astronomers call "lookback studies" of galaxy evolution.[8] In astronomy, such surveys have been at the forefront of promoting the open access to data, assembling large datasets, as well as using computational techniques of data mining and machine learning.[9]

The mosaic of Figure I.1 is part of a dataset of the A2713 field that includes recordings of radiation across the electromagnetic spectrum from radio waves to ultraviolet radiation and x-rays. It was made by MUWAGS, the Multi-Wavelength Galaxy Survey (pseudonym), an international team to which members of the MAMBO group belonged. The MUWAGS collaboration included thirty astronomers from the European Union, Switzerland, the United Kingdom, Canada, the United States, China, and India, including tenured senior scientists, postdoctoral scholars, and PhD students.

The brightest feature in the mosaic is a large dot left (east) of the center. Situated in an apparent square brighter than its environs, and surrounded by a circle of light, four spikes seem to radiate from it. It is one of the nearest objects seen here: an old and cool star in our Milky Way, perhaps only a few hundred light years away, but comparatively bright in the near infrared. In the optical (visible light) exposures used to select this field for study it did not seem particularly noteworthy. For Nadine and her colleagues, it is now a concern as a source of image artifacts.

The mosaic's peculiar outline hints at an even greater worry for the team: the bottom right (southwest) quadrant is missing. Initially, the team's aim was to photograph a square of half a degree by half a degree, covering a bit more than the size of the full moon on the sky. Yet, due to persistent bad weather and limited access to the telescope, team leaders decided to drop a quarter of this field from their observing program. They preferred to "go deeper"

[6] The acronym A2713 stands for the pseudonumeral entry (#2713) in the Abell catalog of galaxy clusters (Abell 1958).

[7] I use the term "deep fields" to include all fields on the sky observed for extragalactic studies with long exposures at multiple wavelengths.

[8] See Blumenberg (1987), Jay (2000), and Cimatti et al. (2019).

[9] See Ivezić et al. (2020) and Ting (2025).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 12 Nov 2026 at 16:31:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.002

(making, processing, and adding more exposures to obtain a more sensitive mosaic) in the other parts, instead of arriving at a "shallow" sum exposure. Three quadrants remain, each corresponding to a set of telescope pointings.

Bad luck with weather at the observatory and its impact on the dataset was an enduring concern for the team. It surfaced again one day at a group meeting, at which Otfried, a senior researcher and Nadine's thesis supervisor, was to report on the status of new observations of A2713 and A2714, another galaxy cluster the team studied. Besides Otfried and Nadine, Otto, and Owen (two other senior astronomers), two PhD students and two postdoctoral scholars were present, as well as myself, an ethnographer. The following exchange occurred at the beginning of this meeting. To capture consequential pragmatic aspects of talk in this transcript, underlining marks emphasis, an ellipsis (…) marks a brief audible pause, hhhh marks an outbreath characteristic of chuckling or giggling, HA-HA-HA marks loud laughter, and double parentheses include verbal descriptions of the talk's delivery.[10] Having received a message from the observatory with the latest update on the observations earlier in the morning, Otfried reports:

Transcript I.1

1 Otfried: Okay … so ehm … maybe it would be good to give a brief … eh … account where we … where we stand at the moment with MAMBO … I want to mention before that … from my side we have … new observations in A2714 … ehh … we had new observations in A2714 already a week ago … or a bit more than a week ago … and tonight hhhh the first observations of A2713 for more than a year hhhhh have happened hhhhh

2 Owen: It is still … it is still there?

3 Otfried: hhh ((*chuckles*))

4 (Group): hhhh HA-HA-HA-HA-HA-HA-HA-HA ((*collective laughter*))

5 Otfried: So … it seems that we make some progress … I am pretty sure …

6 (Group): HA-HA-HA-HA-HA-HA-HA-HA-HA-HA-HA ((*collective laughter continues*))

7 Otto: ((*whispers*)) It has drifted away

8 (Group): HA-HA-HA-HA-HA-ha-ha-ha ((*collective laughter recedes and ends abruptly as Otfried continues to talk*))

9 Otfried: I … I think we have a good chance to get … a very good … to decent data base for A2714 in this year … we have already collected quite a bit

[10] In this book I use elementary conventions of conversation analysis for my transcriptions, simplified from Jefferson's (2004) scheme. See the Appendix for details.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 12 Nov 2026 at 16:31:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.002

These scientists' shared concern for their dataset animates this discussion. Arguably, Owen's question (in line 2) of whether the galaxy cluster A2713 was "still there" is rhetorical. Astronomers do not expect a galaxy cluster to disappear or to "drift away" in the sky, as Otto intimates with his commentary (in line 7), certainly not after a year or so of bad weather. However, despite this being implausible to everyone present, Owen's question elicits collective laughter (lines 4, 6, and 8). Unable to ascribe this laughter to individual participants I have labeled its authors as "(Group)." Otfried can be heard as having invited their laughter. Not only could his choice of words, that the new observations "have happened" (instead of being made), signal irony – his passive voice suggesting that their making was out of the team's control. His chuckling (toward the end of line 1) also opens a slot for Owen to position his question, to which Otfried reacts with a brief, sharp outbreath that continues his chuckling (in line 3).[11] Yet the laughter that follows is contained. It ends abruptly (line 8) and Otfried goes on to give his account on the current state of observations for the project (line 9). One may hear him calling for the group to go on with business as usual, astronomically speaking.

Members of the group probably heard Owen's interjection as a quip on their notoriously bad luck with weather at the observatory. But they may have also heard it as suspending the backgrounded assumption of the night sky's stability, a commonplace for astronomers since antiquity and all too familiar to leisurely stargazers who recognize constellations like Orion and the Big Dipper in the night sky year after year. If galaxies and galaxy clusters were to "drift away," the sky – and astronomical work practices – would literally be "out of order." Anthropologist Mary Douglas writes that a joke "affords the opportunity for realizing that an accepted pattern has no necessity" (1975, 96). The laughter that a joke elicits illuminates a social world held in common with others. Joining the laughter affirms one's membership in it.

That people hold worlds in common with others through sharing classifications and methods of sensemaking is basic to human social life. This is essential for our communications to succeed. Owen's quip points at once to the natural order that astronomers unravel and to the social order of research through which they do so.[12] These two orders go together.[13] Owen's quip suggests that there ought to be shared practices of achieving reference – of

[11] On invited laughter, see Jefferson (1984) and Sacks (1992, vol. 2, 570–575).

[12] As Barnes (1995, 66) notes, we can conceive of social order as the absence of conflict, as the stability of a pattern over time, and as a pattern. Implicated in the other conceptions, the latter is of greatest sociological interest. Following Garfinkel (1956, 1967), I shall consider social order not as a given, but examine how participants (members) constitute, recognize, and maintain it.

[13] Garfinkel (1967, vii) remarks that "in doing sociology, lay and professional, every reference to the 'real world,' even where the reference is to physical or biological events, is a reference to the organized activities of everyday life."

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 12 Nov 2026 at 16:31:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.002

finding the galaxy cluster again – which themselves are unproblematic to these researchers. The sudden end of laughter (in line 8) may well mark that these researchers cannot afford to be skeptics. Their work, and much research with large datasets, inhabits a space of nonskepticism.[14]

If this scene thus leads us to ponder the metaphysical and communicational foundations of these researchers' data production, another demonstrates how they seek to enable absent others to learn from their processed data. When Nadine and her colleagues pondered the weather at Calar Alto the team was also busy assembling its first public release of their A2713 dataset. It was not to contain Nadine's near-infrared observations, but the team's core data comprising of visible-light photographs taken with a telescope in Chile, high-resolution exposures taken with the Hubble Space Telescope, and a mid-infrared map produced with the Spitzer spacecraft, along with a large object catalog – a table of measured and estimated properties of galaxies in this field –, based on these data. An early catalog version included the output of diverse algorithms applied to images like Figure I.1 – positions, measured radiation fluxes, photometric redshifts, error estimates and so on –, merged into a single large table with ca. 88.000 lines (one per detected object) and about 700 columns (one for each measured or estimated parameter). At a collaboration meeting it was cut down to 200 columns, a more manageable size for users beyond the team.

In the team's work, this table became a remarkable thing. To be published along with it, a journal article was to describe the catalog entries and how they were made. MUWAGS team members worried that, as with so many manuals we encounter in all walks of life, users may ignore these instructions or not read them properly, only to blame the MUWAGS team for mistaken uses thereafter. But there was yet another concern about the table: that with any reasonable amount of work its description would be incomplete. Peter, a postdoctoral scholar and MUWAGS member, explained to me:

Transcript I.2

How should data users know this all? You as a team cannot put all your complex shared knowledge into the documentation. You forget things. You take stuff that you know and work with for five years as a given. You forget to include it in the documentation. That happens all the time. (…) Perhaps you could write a massive project blog or a website to explain all that could go wrong with using the data … but we are probably not even aware ourselves of what we would have to describe since this is a years-long trained … but perhaps already intuitive … process.

[14] I thank Douglas Macbeth for this formulation. Like other scientists and everyone else, astronomers cannot meaningfully question presumptions foundational to their work from within their practices. Much of Wittgenstein's (1969) *On Certainty* explores this point. See also Pollner (1987) and Gebauer (2009).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 12 Nov 2026 at 16:31:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.002

Given how hard it is to convey the team's knowledge of their data by describing it, Peter and his colleagues would prefer the catalog itself to encode it. And indeed, catalog entries contain and conserve, somehow, parts of the team's collective knowledge of their data, "frozen in" for future uses and users. Notably, several of the table entries are likelihoods – such as those resulting from the probabilistic classification algorithms used to estimate whether any specific dot or blob in Figure I.1 is a star, a galaxy, or a quasar (a luminous and apparently compact extragalactic object). It is through such likelihoods that the team communicates its beliefs about its data to users, who can employ catalog entries to address new questions, generate further likelihoods, and change their beliefs about the galaxy cluster, its member galaxies, other objects of this kind, the distribution of cosmic dark matter, and so on.[15] To change one's beliefs means to learn. Computation and human learning thus are entangled and the team's challenge is: "How can we share our data so that others can use it to update their beliefs?"

The photographic mosaic and these scenes of its making and publication illustrate how data have become peculiar stuff that scientists use to learn about nature, instruct students, build connections, transmit knowledge, grow trust, and evaluate the work of others.[16] The mosaic of Figure I.1 is one image of the night sky made for scientific analyses, a recording of radiation from deep space. But it is also a deposit of social relations, metaphysical assumptions, and economic considerations.[17] First, the recording medium is central to these astronomers' work. They can add digital exposures to form a single image in which faint signals rise over the noise, becoming detectable and measurable.[18] Second, that these exposures, taken at different times, can be added to form an image is contingent on the night sky's stability – a key aspect of astronomy's natural order. (Other sciences engage other natural orders.) Third, data-rich research and the training of junior scientists go together. Indeed, much scientific work is done in instructional settings, but students learn also by participating in

[15] Here astronomers use "belief" not as something hidden in a private mind (cf. Ryle 1949), but as in Bayesian inference, where it stands for the prior distribution as an estimate of the probability of a hypothesis. Bayesian inference is about updating prior beliefs with likelihoods (Jaynes 2003; MacKay 2003; Sivia 2006). I describe these catalog entries as likelihoods for simplicity. Many are in fact posterior probabilities that cannot be simply converted to likelihoods.

[16] By writing of "stuff" I emphasize the probing and open-ended nature of my inquiry in which I do not know in advance all that matters to the scientists and technicians whose work I witnessed.

[17] For the notion of an image as a "deposit of social relations" see Baxandall (1972).

[18] This additivity is one of the "affordances" (Gibson 1977) of digital imaging in astronomy (cf. Chapter 1).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 12 Nov 2026 at 16:31:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.002

group meetings and other gatherings, such as when they make sense of a quip like Owen's. And they learn on their own and with other students by encountering data, computer code, diagrams, models, and phenomena.

By implication, fourth, a lack of data threatens the group's social organization as a unit of research and training. As with many collaborations in astronomy, MUWAGS did not plan an instrument, build it, and use it to produce data. Instead, its members wrote a series of proposals and were lucky, being granted observing time at major observatories. For a certain period of exclusive proprietary use, the data they obtained were their epistemic and social "stuff," with which they established connections and collaborations.[19] Fifth, for Nadine the lack of data due to bad weather turned into professional and personal trouble. "Her" deep field's depth – variously expressed as integration time or the magnitude of the faintest detectable signal – was her enduring concern. One night at the observatory, when yet another of her observing runs was "clouded out," Nadine worried that "a deep field that does not go deep is not a deep field." With this truism she arguably signaled to me that her aspiration to membership in the community of deep field astronomers was at stake. Contributing to the production of a dataset was a key to her participation and belonging in this community.

The MAMBO team's exposures contributed to the MUWAGS dataset. Its members' give and take of data – as well as the mosaic's missing corner (Figure I.1) – suggests that, sixth, the "economic" is intertwined with the "epistemic" and the "social." These three cannot be properly separated. (And there is always politics as well.) Seventh, issues of knowledge and proper use reach beyond the Heidelberg team and the MUWAGS collaboration toward users of their data, who may hold the MUWAGS team accountable for their work.

There is also a lesson for methodology here. The pervasiveness of instruction in science is a boon for ethnographers who can witness researchers explicating what they usually take for granted. With students being common onlookers to laboratory and data analysis work, there is a place at their side for ethnographers to gain insights into its material and social orders, standards of judgment, background assumptions, and accounting practices of such work. Much of this is lost in studies based on interviews and documents.

[19] The competitive use of shared facilities is not unique to astronomy, as studies of particle accelerators, research ships, and spacecraft demonstrate (Traweek 1988; Law and Akrich 1994; Doing 2009; Vertesi 2015, 2020; Goodwin 2018).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 12 Nov 2026 at 16:31:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.002

We have a lot to learn even from brief moments of interaction. Take Owen's question, for example. The problem of how to interpret it was "thrown back" to participants, whose laughter made it recognizable as a pun.[20] With their response the analyst's position shifts from social scientists to participants in the field. The first to analyze Owen's utterance and recognize it as humor is not the ethnographer – everyone at the meeting arguably did so. As Douglas Macbeth (2007, 200) puts it, "ordinary cultural members are the first analysts on the scene." And so it is not only with laughter but with talk in general. Attending to what is witnessable rather than conjecturing about what goes on in people's minds, not seeking to pass judgments on people's utterances and actions using external standards, as well as not grafting concepts onto observations quickly, means to adopt a naturalistic attitude. This is the approach I take in this book.

I.2 Data-centrism Is a Social Process

The work of the Heidelberg group is but one example of how data are central to social relations and social interactions of scientists and technicians, students and their mentors, teams that collaborate and compete. In their professional lives, scientific data are epistemic and social things. Data are also entangled in economic and political relations, but *How Data Need People* considers the social as foundational, not least because economic and political action also is, at its core, accomplished socially. There is a data-centrism here in the sense that making, using, and publishing data causes humans to interact and engage with each other in specifiable ways. Making and using scientific data necessarily involves people's evaluative tendencies, and, as such, their concerns with their actions being adequately understood by others, with being held accountable for the products of their work, with referring to the same world together, and with becoming – and remaining – adequate members of professional communities.

Adopting this viewpoint helps us to formulate an ethnographic, interactional, practice-focused complement to a rich literature on the production, use, and management of scientific data. When sociologists first began to study scientific work ethnographically, they challenged portraits of science that philosophers had sketched and produced accounts of the (social) construction of facts. Data were a topic of their descriptions, but not a central concern. Thus, in one of the first ethnographies of laboratory work in

[20] See Sacks et al. (1974, 728). The formulation "thrown back" is due to Lindwall et al. (2015, 144).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 12 Nov 2026 at 16:31:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.002

science, sociologist Karin Knorr Cetina (1981) noticed that scientists and technicians used an array of machines to generate and fixate a variety of records and traces on paper, photographic film, and other media. But as these marks were subsequently processed to "manufacture" scientific facts, few of these records seemed to leave the laboratory – unlike the arguments they supported. Bruno Latour and Steve Woolgar (1986, 45–53) called the marks that laboratory machines produced "inscriptions." In their sketch of a biochemistry lab as a site of production, Latour and Woolgar (1986, 46) depict how energy, chemicals, animals, as well as mail and telephone connections enter the laboratory. What leaves it are "ARTICLES" (capitals in original) – presumably containing arguments and newly constructed facts –, but not "inscriptions" or "data."[21]

When these studies appeared in the 1980s, online databases began to transform research in many disciplines, first perhaps in molecular biology, mobilizing data beyond individual laboratories like never before. Stephen Hilgartner (1995) argues that such databases constitute a novel regime of science communication, impact research practices as well as institutional arrangements, and reshape the boundaries of what counts as public and private and possibly even the contents of knowledge. Nowadays, online databases are common and widely used, from genetics to the environmental and climate sciences, astronomy, and linguistics.[22]

With data thus mobilized, social scientists have used notions like communities of practice, moral economies, and gift exchange to capture the social and moral dimension of using, sharing, and publishing data. By doing so they hint at the potential of anthropological and sociological analyses. Thus, Jeremy Birnholtz and Matthew Bietz (2003) argue that data are implicated in scientists' social worlds and communities of practice through ownership and access. Bruno Strasser (2019) emphasizes the diversity of scientists' moral economies in the sharing and withholding of genetic-sequencing data. Christine Borgman and her coauthors recognize data as a collaboration's "glue" (Borgman et al. 2012, 485). Echoing Marcel

[21] Bogen and Woodward (1988) define data as local evidence for nonlocal phenomena. Leonelli (2016, 77) defines scientific data inclusively as "any product of research activities (…) that is collected, stored, and disseminated *in order to be used as evidence for knowledge claims*" (emphasis in original). Bokulich and Parker (2021) restrict their definition to records that are separate from the system or phenomenon under study. By following Hacking (1992a, 48) and considering only the products of data generators (typically forms of writing or inscription, broadly conceived) as data, my view is yet narrower and more specific.

[22] See Hine (2008), Leonelli (2016), Hilgartner (2017), and Strasser (2019) on genetics; Edwards (2010) and Borgman (2015) on environmental sciences; McCray (2014, 2017) on astronomy; and MacWhinney (2025) on linguistics.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 12 Nov 2026 at 16:31:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.002

Mauss (2016 [1923–1924]), they identify a gift culture in which datasets are assets that can be traded with other researchers, used as leverage in collaborations, and brought as dowry (Wallis et al. 2013). If data are released, Borgman (2015, 217) argues, they can lose their value as "assets" to "barter."

But owning, accessing, and publishing them as resources is only a part of how data are implicated in the social, economic, and political lives of scientists and technicians. Philosopher Sabina Leonelli connects an understanding of data as "tools for communication" (2016, 69) with an examination of their epistemic function and economic value. Drawing on interviews with database curators, Leonelli argues that data-centrism marks a shift away from considering data as single-use products of the research process toward recognizing "efforts to mobilize, integrate, and visualize data (…) as contributions to discovery in their own right" (Leonelli 2016, 2). In her view, it "consists of a normative vision of how scientific knowledge should be produced in order for the research process to be efficient and trustworthy" (Leonelli 2016, 197). Good data, it seems, need curators who, literally, care for them.[23]

Studies of information infrastructures also highlight how important human action and skill are for enabling the reuse of data. Christine Borgman (2015, 4) argues that data "exist within knowledge infrastructure – an ecology of people, practices, technologies, institutions, material objects, and relationships." Paul Edwards' (2010) historical study of research about climate change argues that this phenomenon becomes recognizable to scientists only through the stable background that the knowledge infrastructures of climate science provide. These include networks for standardizing and calibrating instruments as well as practices of measuring, distributing, and repairing meteorological data. Edwards and his coauthors (2011) note that where proper infrastructures are not established, users often understand others' data only through elaborate communications with data makers that they call "data friction." They recognize the parallel of this process to everyday social interaction, in which "common ground" – shared understandings fundamental to mutual sensemaking – is typically provisional and contingent. For data to be usable, people need to do this work, but here, as well, these people's commitments and accountabilities remain largely unspecified.[24]

[23] See also Pinel et al. (2020) and Pinel and Svendsen (2024).

[24] On "common ground," see Lewis (1969), Clark (1996), and Enfield (2013). Scientists widely recognize the importance of infrastructures for data-rich research (cf. Wynholds et al. 2011; Blanton et al. 2023). Nevertheless, as I focus on practices of making, using, and preparing data for publication,

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 12 Nov 2026 at 16:31:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.002

We can probe yet further into the socialness of data-rich science and find it also in how scientists draw inferences from data by computational means. Linguistic anthropologist Paul Kockelman (2017, 2025) proposes a Bayesian anthropology that treats reasoning as probabilistic and computational. He seeks to align this approach with the semiotics of Charles Sanders Peirce, who argued that meaning is generated in chain-like processes of semiosis in which signs instigate the production of new interpretants (objects of interpretation). However, Kockelman does not address how embodied, evaluative, social humans participate in this process and strive to be accountable to other researchers. This is what a naturalistic anthropological approach to data-rich science can examine.

I.3 An Anthropological Perspective

An anthropological approach, as I understand it here, puts people – embodied, cognizant, social, and cultural human beings – and their practices at the center of attention. It begins neither with ready-made concepts nor with individuals and "private ratiocination" (Douglas 1992, 240), but with public understandings and shared practices. That we are social before we are individual is a viewpoint owed to Émile Durkheim, a founder of sociology and anthropology.[25] It is a sense of anthropology endorsed by philosopher Ludwig Wittgenstein, who in his late work observed shared social practices and wondered how knowing something is a witnessable ability.[26] Wittgenstein insists that learning to do what others do by attending to what they make explicit and what they take for granted does not require access to their private thoughts but can be achieved through participating in a shared form of life. Thus understood, what novices in science (and elsewhere) do bears some resemblance with an ethnographer's practice of participant observation.

Such an anthropological approach ought to focus on communication and do so in a fine-grained way that is alert to social action and interaction. Language is arguably our principal means to coordinate social actions, to build and maintain social relations, to commit ourselves to social projects, and to hold each other accountable. There is a tyranny of accountability in

but not archiving or retrieving them, infrastructures play a relatively small role in my account. See Genova (2018) and Borgman and Wofford (2021) on astronomy's data infrastructures.

[25] Durkheim makes this point in *Elementary Forms of the Religious Life* (1915 [1912]). Interaction-minded scholars like Mead (1934) and Blumer (1969) share this viewpoint.

[26] Consider, as an example, paragraphs 143 to 184, on understanding and ability, in Wittgenstein's (2009 [1953]) *Philosophical Investigations*.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 12 Nov 2026 at 16:31:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.002

social life, "an ever-present possibility of being noticed, praised, blamed, questioned, called out, and judged" (Enfield and Sidnell 2022, 21). This accountability is exercised largely through uses of language.[27] If an ethnographic study can sketch a "model of the human" in data-rich science, social accounting practices are bound to be one of its foundations.[28] But such an anthropological approach, and such a model, must also acknowledge the biographical nature of participants' lives.[29] Membership in a community and culture, credibility, and reputation are biographical matters, and so are assessments of genealogy ("Who was her supervisor?") and expertise, such as when someone's "track record" is examined (Collins and Evans 2007). Biographies are repositories of trustworthiness.

When anthropologists explore an unknown domain, they are often wary that diverse actors, agencies, and objects, not all human, may matter to a social setting. They cannot know in advance which forms social life may take, nor if these actors and agencies are members of a culture or community. Compared with terms like community, moral economy, or culture, the notion of "sociality" is more inclusive and open-ended, less laden with presumptions about the workings of social relations, interactions, and practices. Some anthropologists use sociality as a term that is not in need of a definition (Graeber 2011; Wilf 2013; Miller 2015). Others have defined it as the "art of living together" (Lee 1927), as "the capacity to be social" (Schick 1984), as "referring to the creating and maintaining of relationships" (Strathern 1988, 13), as a "condition of social co-presence" (Chau 2006, 147), as "the character of social interaction that underpins social life" (Enfield and Levinson 2006, 2), and as the "range of possibilities for social coordination with others" (Ochs and Solomon 2010, 69).[30]

[27] See Evans-Pritchard (1937), Douglas (1980), Levinson (1983), and Enfield and Sidnell (2017).

[28] Such a model may be a caricature, but it could turn out to be a useful heuristic for the social study of future developments in science. The study of accounting practices can also complement and develop pragmatic philosophical views of what "epistemic agents" (Chang 2011, 2022) are like. After all, epistemic agents are unavoidably also social agents, and as such users and subjects of accounting practices. It could also contribute to a rapprochement of the history and philosophy of science with the sociology of science, which some critics (like Daston 2009) have regrettably reduced to approaches like Actor-Network Theory and the Sociology of Scientific Knowledge, to the neglect of alternatives (Kinzel 2012), including the naturalistic and interactional viewpoint that I adopt. This would also benefit approaches like cognitive ethnography (Nersessian and MacLeod 2022) and discipline-specific studies (e.g., Boyd et al. 2023), which, so far, lack interactional views of human reasoners and actors. See Veigl and Currie (2025) for the use of social science methods in the philosophy of science.

[29] Note that, for Gell (1998, 10), anthropology "tends to focus on the 'act' in the context of the 'life' – or more precisely, the 'stage of life' – of the agent. The fundamental periodicity of anthropology is the life cycle."

[30] Notions of sociality are also popular with philosophers (Tuomela 2007; Nguyen 2020) and behavioral scientists (Ward and Webster 2016).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 12 Nov 2026 at 16:31:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.002

In this book I examine data-centric socialities as forms of social coordination unfolding in the making, use, publication, and reuse of digital scientific data. My notion of data-centrism – that making, using, and publishing data causes people to interact and engage in specifiable ways – is inspired by Georg Simmel, who early in the twentieth century proposed the study of sociations (*Vergesellschaftungen*) as the proper object of sociology. Simmel noticed that "neither hunger nor love, neither work nor religiosity, neither technology nor the functions and results of intelligence (…) are themselves sociation. The latter is manifested only when individuals' isolation is transformed into forms of togetherness and for-each-otherness that count as interaction" (Simmel 1992 [1908], 18–19, my translation). He continues that the "specific subject matter and the form of sociation are always a unit, since a social form cannot have an existence when dissolved from any content, as much as a spatial form cannot exist without the matter whose shape it is" (19). Simmel sensed that "minor forms of sociation" (34), which ought to be studied "microscopically" (34), shape society at large. He distinguished forms of sociation at different scales, from fleeting encounters to lifelong and generational forms of belonging to families, medieval guilds, and nation-states.[31]

Elsewhere, Simmel (2004 [1900]) pondered the effects of money, a quantitative medium like many forms of data, on forms of sociation.[32] Simmel argued that, unlike barter, uses of money always introduce the community as a "third party" to an exchange. Money, he writes, therefore is a "claim upon society" (Simmel 2004 [1900], 176–177). Thus inspired, I consider a variety of data-centric socialities that involve scientists and technicians, students and their mentors, teams that collaborate and compete, participants of hack weeks, as well as the technologies and media that they use together.

Inspired by Simmel, we benefit from a change in perspective. For good reasons a growing body of studies has probed into pressing ethical issues of contemporary data uses in science and society.[33] But when we attend to interactions of data makers and data users ethnographically, a subtle ethical landscape comes into view that is consequential for those affected, yet unrecognized by existing studies. Scientists and technicians emerge as social inquirers themselves (Chapters 2, 3, 7, 8, and Outlook). Combining

[31] On Simmel as an inspiration, but not a model, for empirical investigations of social interaction, see Bergmann (2011).

[32] See also Dodd (2014).

[33] See, for example, Floridi and Taddeo (2016), Mayernik (2017), White et al. (2022), Poole (2023), and Padmapriya and Parthasarathy (2024).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 12 Nov 2026 at 16:31:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.002

data from different sources for analysis – a common, often challenging problem of data-rich science – becomes a window into a discipline's community and culture (Chapter 3). We discover how diagrams are sites for "cultivating data" (Chapter 4) and domains for practicing mundane reason (Chapter 5), and we learn how scientists use collaborations as organizational experiments to manage their work with large datasets (Chapter 6). Probing into the "inner conversations" among members of a research team gives us novel insights into the normative expectations of "open science" (Chapter 7). And we learn that scientists commonly presume a "reciprocity of perspectives" (Schütz 1962, 1964), assuming that data users act like themselves, as they seek to encode their knowledge in a dataset and design it for others to use it properly (Chapter 8).

As suggested, my approach is naturalistic. I attend to what I can witness as an ethnographer. I accept Edwin Hutchins and Brian Hazlehurst's (1995, 54) "no telepathy assumption": "No mind can influence another except via mediating structure." I refrain from conjecturing about what goes on in people's minds and do not seek to pass judgments on their utterances and actions using external standards. Instead, I aim to make sense of data-centric socialities from the viewpoints of their participants: scientists and technicians, novices and old hands.

My perspective is aligned in many ways with ethnomethodology, a sociological approach to the study of practical actions and practical reasoning. Ethnomethodology is not a method and it is not ethnography. Ethnography is the empirical study of people living and acting in their life worlds, undertaken by observing and participating in these settings.[34] Ethnomethodology is the study of the methodical ways in which people make sense of ordinary and professional activities and do so in ways that others can recognize. Following Harold Garfinkel (1967), who modified advice that Durkheim (1982 [1895]) gave in *Rules of Sociological Method*, ethnomethodological investigations are guided by the heuristic to treat social facts as accomplishments: "Where others might see 'things,' 'givens' or 'facts of life', the ethnomethodologist sees (or attempts to see) process: the process through which the perceivedly stable features of socially organized environments are continually created and sustained" (Pollner 1974a, 27). As Michael Lynch puts it, "[w]hat is at stake is not the

[34] In my view, ethnography is not simply a method either, but an approach that is imbued with a particular epistemology and ontology, such as by situating "language deeply and inextricably in social life" (Blommaert and Jie 2020, 7), by attending to the viewpoints of participating humans (Agar 1995; Fabian 1995; Hymes 1996), and by examining their sensemaking practices. Uses of numbers, measurements, and diagrams are, likewise, situated in the social life of data-rich science.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 12 Nov 2026 at 16:31:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.002

theoretical problem of order, but the substantive *production* of order on singular occasions" (2001, 140; emphasis in original). This order is typically brought about sequentially and it can often be analyzed as such. Consider again the conversation of Otfried, Owen, and their colleagues that I witnessed, recorded, and transcribed (Transcript I.1). Following Otfried's complaint about weather at the observatory, Owen posed his question and participants analyzed it "on the fly" to act thereupon: they laughed. Thus, ethnography can yield materials for an ethnomethodological inquiry.

Ethnomethodologists and conversation analysts argue that we analyze each other's actions all the time and that we can gain insights into participants' understandings by attending closely to how they respond to each other, as it reveals the order that they create, recognize, sustain, and take for granted.[35] In fact, we can conceive of any social actor as a lay sociologist or anthropologist, who is "an inquirer into the practical circumstances that confront the member going about the business of everyday life, an ethnographer of its culture, a methodologist separating truth from falsity, fact from fancy, valid from invalid inferences, reality from fantasy" (Hester and Eglin 2017, 200).

If anthropologists have emphasized the importance of social accountability for social order and mutual sensemaking, a notion of accountability is, likewise, central to ethnomethodology. There it refers to the self-organization of social settings by its members who "make its properties as an organized environment of practical activities detectable, countable, recordable, reportable, tell-a-story-aboutable, analyzable – in short, *accountable*" (Garfinkel 1967, 33; emphasis in original) to each other. There are two aspects of this accountability: producing the order as well as making it witnessably available so that it can be reproduced. Achieved through "situated practices of looking-and-telling" (Garfinkel 1967, 1), this communicative work establishes the participants' trust and the moral order of their interaction.[36]

Note the important distinction between accountability as a concern with being held to account (a reflective stance) and ethnomethodological

[35] Thus understood, human sociality is characterized by processes of mutual, methodical sensemaking. Talk-in-interaction in copresence may be its foundational site. Schegloff (2006, 70–71) summarizes an enormous body of scholarship when he writes that "[p]eople talk in turns, which compose orderly sequences through which courses of action are developed; they deal with transient problems of speaking, hearing or understanding the talk and reset the interaction on its course; they organize themselves so as to allow stories to be told; they fill out occasions of interaction from approaches and greetings through to closure, and part in an orderly way." See also Sacks (1992) and Sidnell (2010).

[36] See Garfinkel (1963); Lempert (2013).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 12 Nov 2026 at 16:31:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.002

accountability as situated within lived practice (and thus often nonreflective), shaping how actions are organized. The former requires language for becoming socially relevant, whereas the latter may include uses of language but is commonly prelinguistic or even nonlinguistic. I consider accountability-as-responsibility as a second-order phenomenon that commonly presumes ethnomethodological accountability. I shall refer to "accounting practices" inclusively as the actions oriented to the achievement of either kind of accountability.

As ethnomethodologists understand it, reflexivity is a characteristic feature of social interaction, during which participants continually reassess earlier observations, utterances, and interpretations. Always understood as temporal and sequential, ethnomethodological reflexivity is different, for example, from the postmodern concern of ethnographers about their role in fieldwork. For ethnomethodologists, reflexivity is implicated in the phenomenon of accountability.[37]

I.4 The Social, the Numerical, and the Computational

As a "social communication technology" (Dor 2015) language may be our principal means to coordinate social actions, to commit to social projects, to hold each other accountable, and to build and maintain social relations.[38] But numbers, computations, diagrams, and pieces of computer code are other means to do so. Wittgenstein (1978) noticed that confusion would result if we were not to agree routinely on the result of simple calculations, such as that 25 × 25 is 625. Making such calculations and routinely agreeing on their outcome is, for him, an anthropological fact. This numerical accountability is a social accountability.[39] Given that most

[37] See Heritage (1984) and Lynch (1993). Barnes et al. (1996, 205fn4) describe ethnomethodology as "a means of studying realisms," and this suggestion, I believe, makes it rewarding to bring it into conversation with recent work in the philosophy of science that strives for "realistic" views of (scientific) realism (Chang 2022). If anthropologist Tim Ingold (2018, 2) is right to claim that "anthropology is philosophy with the people in" then one might as well wonder what a "(scientific) realism with the people in" would look like. Sharrock and Anderson (1991, 74) argued long ago that the question of "whether scientists are right in their 'realistic' construal of their achievements gives way to the question of whether the scientists' sense of the reality of the phenomena they deal with has in fact been identified at all." Ethnomethodological studies are valuable for this purpose, but see Hoeppe (2023) for some reservations about ethnomethodological studies of scientific practice.

[38] However, see Leroi-Gourhan (1993 [1964]), Gundaker (1998), and Rutherford (2025) on uses of touch, gestures, inscriptions, and graphisms.

[39] Rawls (2002), Pollner (2012), vom Lehn (2016), and Koschmann (2019) speculate that Harold Garfinkel's notion of accountability may have been inspired by the bookkeeping he witnessed at his parents' furniture company.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 12 Nov 2026 at 16:31:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.002

scientific measurements are numbers, and digital data are intrinsically numerical, this medial microstructure must be consequential for data-centric socialities (Chapter 5).

Digital images like Figure I.1 are two-dimensional arrays of numbers. They are used for diverse analyses and measurements in astronomy and other sciences. Their numerical accountability reaches beyond dimensionless single numerical values and extends to one, two, or even higher dimensions (cf. Chapters 1 and 5). How digital photographic exposures are formatted is essential to what astronomers can do with them, such as adding, subtracting, or dividing exposures. It is also essential to this work's social organization – such as when several smaller teams combine data taken at different telescopes into a joint dataset, as was the case with the MUWAGS collaboration. The call is, therefore, to include the material properties and medial formats of data in an anthropological analysis.[40]

In pondering how data, as media, shape social forms, we can take inspiration from anthropological views of money use throughout human history. Keith Hart (2001, 2016) observes that, as a symbolic medium, money not only reduces the value of diverse entities to a single scale and conveys information more easily. It is also a "store of memory linking individuals to their various communities, a kind of memory bank (…) and thus a source of identity" (Hart 2016, 7).[41] Nadine's quest for membership in the community of astronomical deep field researchers by contributing to the production of a dataset resonates with this observation, as it does with Simmel's view of money as a "claim upon society" (Simmel 2004 [1900], 176–177).

Many anthropologists take a negative view of money, highlighting its coercive function in exploitative and unequal economic relations, but Hart recognizes in money a medium that humans can shape to their needs, a medium that also has redemptive features: "Money is how we learn to be truly human" (2016, 11). Hart's argument for a "human economy" from which no one is excluded resonates with demands for the open access to data and other resources in the sciences. This aspiration arguably guides many hackathons and hack weeks – meetings at which participants work together with public datasets on shared problems, teaching each other methods and hacks along the way (Chapter 6).

[40] Diverse material artifacts can mediate human action and knowledge (Goody 1977; Vygotsky 1978; Marx 1990 [1867]; Hutchins 1995; Klein 2003; Latour 2005; Goodwin 2013, 2018; Madianou and Miller 2013). See Munro (2001) for a view on how numbers mediate social life, an important topic for accounting studies and the sociology of science (Heintz 2007; Mennicken and Vollmer 2007).

[41] See also Dodd (2014) and Maurer (2021).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 12 Nov 2026 at 16:31:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.002

I.5 Data Economies

Christine Borgman argues that datasets are assets that can be traded with other researchers, used as leverage in collaborations, and brought in as dowry. Released data, Borgman (2015, 217) reports, lose their value as "assets" to "barter." She adds that data may also be regarded as "liabilities" (2015, 217) due to the costs and efforts of storing and managing them or because of legal implications, such as when guarding rights to anonymity and privacy. These descriptions focus on concerns over ownership and access, but do not address how data would be used.

As Hart (2001) points out, we are increasingly imagining living together as an economy. Many astronomers are acutely aware of their work's economic dimension and their references to it go along with a shift toward data-centrism. I was intrigued by how pervasively my interlocutors used a vocabulary borrowed from capitalist economy to describe the organization of contemporary observing. Astronomers commonly refer to data as commodities, talk about where best (on the sky) to invest a block of telescope time, and describe observing projects as enterprises oriented to making profits. Although some astronomers had referred to data as commodities before,[42] this metaphor arguably gained traction in the early 2000s. By then several major research institutes had cut long-term contracts with observatories and instead joined collaborative projects where they would share the financial costs of building and operating instruments for access to their data products.[43]

As a notion to describe the access to and use rights of data, "economy" is both a member's term and an analyst's term. But analysts like Borgman and participants use it only to highlight specific aspects and not to describe the workings of an entire discipline. Considering a domain of practice as an economy might suggest that an exchange principle is basic to all its activities, or that all these operate according to a capitalist logic. This can be misleading.[44]

I.6 Scales of Organizing: Collaboration, Friendship, Hack Weeks

Doing science with large datasets involves diverse actors and relations. It is differentiated. To get at this differentiation, we need to develop

[42] For example, Léna (1988, 1).
[43] Hoeppe (2018a). The ALLSKY survey described in Chapter 7 was such a project.
[44] See Chapter 7 on the notion of "moral economy."

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 12 Nov 2026 at 16:31:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.002

both a more fine-grained vocabulary and an understanding of diverse scales of sociality in terms of the number of participants and the complexity of technology involved, as well as the temporality that unfolds. Contemporary astronomers are not liberal subjects who own data privately and permanently. Large amounts of observing time at public observatories are made by, and given to, collectives only. Structured and often hierarchical, collaborations are dominant in data-rich science and central to much of what I observed. Collaborations have lifetimes from a few years to decades. Many of the long-lasting collaborations described in the literature are formal organizations marked by their privileged or even exclusive access to data-generating facilities.[45] When focusing on facility users, I find that many collaborations are "organizational experiments" (Sharrock 2011, 29) that draw on diverse resources to order their affairs, often in improvisational ways (Chapters 6, 7, and 8).

Several large data-centric collaborations in astronomy have developed from the friendship of actors in influential organizational roles. Thus, "old boys'" networks (McDonald 2011) certainly matter. But friendship, as a form of social relation characteristic of settings in which other rights and obligations are suspended, is important more generally in a field in which shorter-lived and often fluid social forms emerge. In astronomy's diverse ecology of data production, where access to many, but not all, facilities is nominally open to all researchers, networks of friends and ties of academic genealogy provide diverse ways of "mutual aid" in organizing access to data, assembling the expertise to analyze it, and establishing coauthorship (Chapter 6). Human work lives are these ties' characteristic timescales.

Compared to long-lasting formal collaborations and the affective ties of friendship, hackathons are fleetingly short. Rarely lasting more than the five days of a "hack week" (a common format), hackathons in astronomy are typically marked by a diverse group of participants who work ostensibly free from hierarchies on newly released datasets, with open software, and on shared problems. Hackathons are thus a social form contingent on openness. Instruction, learning, and building new relations are their goals.[46] While hackathons may be brief, participants' relations are often more enduring.

An interactional approach to these scales of sociality could be reductionist and trace all social structures to patterns of interaction. But, in adopting it, analysts would ignore the diverse contexts of accountability and

[45] Knorr Cetina (1999), Shrum et al. (2007), Vertesi (2015, 2020), and Roy (2024).
[46] Huppenkothen et al. (2018) and Falk et al. (2024).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 12 Nov 2026 at 16:31:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.002

accounting practices in which members engage. As Egon Bittner (1965) noted long ago, people use "organizational constructs" – understandings of how to act as a member of an organization – as resources for achieving, and mutually making sense of, intelligible actions.[47] Inspired by Bittner I examine how this inherent reflexivity shapes courses of practical action. I find that scientists use diverse resources, including medial and epistemic orders as well as a dataset's projected coherence, for organizing collaborative work (Chapter 6).

I.7 Learning from Instruction

As mentioned, the pervasiveness of instruction in science is fortunate for ethnographers, as there is already a place for them available at the side of students, inviting insights into backgrounded assumptions and into how practices unfold in real time. Thus, we can identify problems of data-rich science, consider how they are "staffed" with people in a specific case, and witness ethnographically how they are managed.[48] In this way, scientists' social, epistemic, political, and economic accountabilities become unavoidably part of our analysis. Yet more intriguing is a convergence of learners' practices that includes junior scientists and ethnographers, a theme that I shall examine in several contexts (Chapters 2, 3, 7, 8, and Outlook).

Data-centric socialities are inevitably media-centric socialities, and thus my exploration of instruction begins in Chapter 1 with an inquiry into what scientists can do with digital media that they could not do before. Setting out from a lecture on image processing to undergraduate students, it traces astronomers' understandings of digital data's affordances. It argues that the introduction of charge-coupled devices in the 1980s provided solutions to a set of practical problems that astronomers had formulated with increasing clarity since the 1950s. Subsequently, new organizational possibilities for astronomical research emerged. These include mobilizing data beyond local contexts, rendering abstract time as an object of management, sharing data as nonrivalrous goods, assessing others' work remotely, and building new forms of collaboration – elements of a novel medial middle ground for data-rich science.

[47] Weick (1979, 1995), Anderson et al. (1989), and Anderson and Sharrock (2018).

[48] My approach to consider how problems are "staffed" is inspired by Garfinkel's examination of variously populated waiting lines as examples of "immortal ordinary society" (2002, 65), but sociological studies of situationalism (e.g., Marres and Stark 2020) are relevant as well.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 12 Nov 2026 at 16:31:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.002

Chapter 2 returns to human sociality and supplements my introduction to elements of ethnomethodology (earlier in this introduction) with a review of how humans are fundamentally evaluative and what the ethical implications of this entail. I examine this by joining graduate students, junior scientists, and technicians as they record data at an astronomical observatory. The "legitimate peripheral participation" in "communities of practice" (Lave and Wenger 1991) has become a catchphrase for the situatedness of apprenticeship. Considering situated interactional engagements of instructors and students can shed light on how this "peripheral participation" succeeds, I show how learning and instruction are situated locally while being subject to accountabilities that extend far beyond any single site. I argue that the visitors' ordinary interactional competence enables much of their learning. The telescope control room's progressive redesign and relocation has modified this experience and facilitated its transfer. But it turns out that technicians' ethical evaluations of researchers do not travel as easily. This is part of an ethical landscape that was invisible to prior studies.

Chapter 3 identifies PhD student training as a curious process in which instruction and the advancement of science go together. It examines how Nadine, the PhD student encountered earlier, was instructed to tackle a common, though often challenging, problem of data-rich science: calibrating a new dataset and combining it with data from a different source for analysis. By following Nadine around over two years as she achieved this goal, we learn how she became a competent member in the community and culture of extragalactic astronomy. Conversely, we gain insights into what makes combining scientific datasets often so challenging. I introduce and adopt Trevor Pinch's (1985) notions of "externality" and "evidential context" as aids for comparing epistemic practices; examine how models are used to calibrate a dataset; study the use of shared, but backgrounded notions about the world; and probe into uses of diagrams and visualizations.

Instruction in data-rich science unfolds not only in face-to-face situations and through user manuals. Ideally, data themselves would instruct their users. How can scientists pass on to potential users what they have learned about their data? Chapter 8 proposes answers to this question by focusing on the medium of data and its social uses. It argues that scientific data do not merely represent information but can be structured and presented to have a pragmatic function oriented to enable users' understanding. It demonstrates this by describing how the MUWAGS collaboration designed its catalog to guide users to self-correct wrong uses and delimit

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 12 Nov 2026 at 16:31:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.002

being held accountable for misuses. Some astronomers argue that catalogs encode their makers' collective knowledge of their data. Chapter 8 examines this claim ethnographically, identifies elements of a pragmatics of data reuse, and ends with a reflection on socio-computational orders – entanglements of the social and the computational in data-rich science.

The MUWAGS team succeeded in making its catalog "more-than-representational" or "more-than-evidential" by projecting its members' own conduct onto imagined catalog users. As such, they assumed a "reciprocity of perspectives" reminiscent of what social theorists following Alfred Schütz (1962) have recognized as a foundational feature of human intersubjectivity. This is not the only instance in which elementary aspects of social interaction and intersubjectivity help us to make sense of data-centric socialities even where scientists do not meet face to face. Other examples are the sequentiality of actions, participants' mutual monitoring, and the reflexive maintenance of inhabiting a shared world, as well as the omnipresence of accounting practices. *How Data Need People* thus examines the uses and limits of interactional understandings when scientists use media in orienting toward others who are commonly absent.

I.8 Learning from Practical Reasoning

Successful instruction often engages learners in experts' practical reasoning. It is instructive to distill elements of this reasoning, not least for making ethnographic insights into one scientific discipline relevant for studies of others. Garfinkel (2022) called for specifying the "discovering sciences of practical action," arguing that scientists "are *investigating* as well as *using* practical actions, reflexively discovering a local organization of practical actions as well as *what* those practical actions disclose, stumble upon, negate, or prove" (Lynch 2022, 10; emphasis in original). Whether novices or old hands, scientists invariably reason "in the midst of things," as Garfinkel (2002, 101, 249, 250) put it. For practical reasoning there is "no time out" (Garfinkel 2002, 118). It is always situated in the here and now, responds to actions, and calls for responses. It is necessarily prospective (looking forward to what is meant to be achieved) and retrospective (informed by what has happened before), open to correction and oriented to various accountabilities – epistemic, social, and otherwise. What scientists do invariably relates to work done by others elsewhere, and is oriented toward writing a report – the long-lasting record of their research.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 12 Nov 2026 at 16:31:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.002

The things in whose midst scientists reason are typically defined by their discipline's subject matter. Geologists study rocks and sediments, botanists study plants, and astronomers study objects "on" the sky. Consequently, research work in astronomy differs from that done in botany or geology. Scientists' reasoning is tied to the materials and tools at hand, including computer code, databases, and data of specific formats.[49] But how does this matter to screenwork's virtuality in data-rich science, which, at first glance, erases the material differences of disciplinary work?

How Data Need People examines how discipline-specific objects and structures become actionable through their mediation. Chapter 5 takes uses of the sky in astronomy as an example. Astronomer David W. Hogg once wrote in a blog post that "[a]ll of astronomy and astrophysics is built on the observation and reobservation of sources on the sky."[50] This apparently self-evident statement came to intrigue me as I witnessed data analyses and listened to conversations such as the one of Otfried, Owen, and their colleagues (Transcript I.1). I noticed that the sky's phenomenal properties – its apparent immutability and the richness of its visible features – pervaded astronomical research and provided infrastructural resources for actions that researchers rarely acknowledge in their publications. The sky (and objects on it) is both a topic of research and a resource for its conduct, as it provides saliences that scientists use alongside existing records for ordering work, diagnosing trouble, and repairing data. It is a resource for relieving data users from trusting data makers. Other sciences, I argue, can dwell on other resources to do so.

Note that this work engages what sociologist Melvin Pollner called "mundane reasoning": recognizing and resolving disjunctive experiences using shared sensemaking practices that draw on the presumption that "reality is coherent, determinate and intersubjectively accessible" (Pollner 1987, 47). Mundane reasoning pervades scientific work with data. Diagrams are one of its resources. Diagrams exhibit data in various stages of processing, from "raw" detector outputs to "science plots" that present measured and derived physical values, such as, say, the masses or

[49] Thomas Kuhn hints at this with his notion of "disciplinary matrix" (1970, 182), a discipline-specific ensemble of equipment, practices, and discourse.

[50] hoggresearch.blogspot.ca/2008/03/budavari-and-szalay.html (accessed July 7, 2025). The formulation "on the sky" is prevalent in astronomers' talk. It acknowledges the "celestial sphere" as a perceived two-dimensional surface and arguably also the mediated nature of astronomical research, where the "sky" is sometimes defined as "a two-dimensional distribution of intensity of electromagnetic radiation" (Léna 1989, 245; cf. Chapter 5).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 12 Nov 2026 at 16:31:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.002

luminosities of stars and galaxies. Scientists rarely talk with each other about their work without also looking at plots or sketches on paper, blackboards, or computer screens.[51] Previous studies have pointed out how diagrams make data, models, and phenomena accessible intersubjectively and available for experimenting, exploring, and operating. I examine how diagrams thus become not only "places of thinking" and discovery,[52] but also spaces in which scientists prune and cultivate datasets in light of epistemic and social accountabilities (Chapters 3, 4, and 5).

Scientists are practical reasoners also when organizing collaboration – inevitable for making, processing, and analyzing large datasets. Chapter 6 examines data-centric collaborations as "organizational experiments" (Sharrock 2011, 29) that draw on diverse resources to order their affairs, often in improvisational ways. These resources include medial formats, epistemic orders, and a team's joint orientation to produce a consistent dataset. Diverse in origin and structure, such data-centric collaborations typically lack features of formal organization, such as organizational charts and legally binding contracts. They often emerge from ties of academic genealogy and friendship. For a certain period of exclusive proprietary use, the data that such teams obtain and make is their epistemic and social "stuff," which members use to establish connections and collaborations.

As practical reasoners, natural scientists are also social inquirers. For many scientists the emergence of "open science" has been a practical problem. Chapter 7 considers how researchers examine the social and moral accountabilities of the open access to data and the tensions that result from them. I noticed that as two collaborations prepared datasets for public release, their members became both inquirers into and actors in what many astronomers refer to as their discipline's culture of open data access. Both teams can be described as groups of practical methodologists who used "inner dialogues" to explore the normative expectations and tensions of this domain, seeking to inhabit proper statuses in it. The chapter argues that examining scientists' understanding of statuses and their achievement offers resources for a refined critique of "open science" that is considerate of the context sensitivity of data production and use. Along the way, it examines some methods of how scientists, as members, are doing ethnography.

[51] See Ochs et al. (1994, 1996) and Alač (2011).
[52] Compare Wittgenstein's (1960, 7) remarks in the *Blue Book* on paper as a place of thinking.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 12 Nov 2026 at 16:31:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.002

I.9 Varieties of Data-Centric Socialities

The episodes examined in this book lead us to discover a data-centrism in contemporary science in the sense that making and using data invites humans to interact and engage in specifiable ways – from moments of instruction to striving for membership in a community and organizing collaborative work. We can find data-centric socialities at many scales. They involve scientists and technicians (Chapter 2), students and their supervisors (Chapters 3, 4, and 5), teams that collaborate and compete (Chapters 4, 6, 7, and 8), makers and users of data (Chapter 8) – and always also the technologies and media that they use together. Never is "the social" simply grafted upon something that is otherwise not social. It is always a constitutive and essential aspect of scientific work with data. Its explorers are not just visiting anthropologists and sociologists, but students, scientists, and technicians themselves (Chapters 2, 3, 7, 8, and Outlook).

As I described earlier, my notion of data-centrism is inspired by Georg Simmel's explorations of the diversity of social forms in the early twentieth century. Probing into a wide array of social forms in his *Sociology* (1992 [1908]), Simmel realized that many forms of sociation await discovery – not least by means of more detailed, "microscopic" studies, but also because new forms keep emerging while others fade away. This, certainly, is to be expected, as well, of data-rich science in the twenty-first century. The chapters that follow offer microscopic examinations of some of its dominant forms of sociality and their resources. While these become visible only at the fine granularity of social interaction and the technical detail of a specific science, the lessons we learn are more general, because uses of social accounting practices are. Sciences other than astronomy may, of course, use other media, engage other disciplinary objects, and use other epistemic orders to organize their work. But they will need to train new members, use diagrams to interpret data, use forms of mundane reasoning, organize collaborations by engaging certain orders, act normatively, transmit knowledge – and engage accounting practices along the way.

I.10 This Study

This study began with eighteen months of ethnographic fieldwork by the MAMBO team – the MUWAGS collaboration – as well as associated and competing researchers in 2007–2010, followed by annual revisits in 2010–2019, additional fieldwork in 2023, and ongoing conversations and exchanges with former team members and other scientists. I witnessed data

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 12 Nov 2026 at 16:31:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.002

analysis work, instructional meetings, team meetings, and teleconferences; conducted interviews; accessed emails; read publications relevant to the team's work; and assisted in a small part of their research. I recorded oral histories and biographical interviews and studied documents archived by participants. I also attended collaboration meetings, seminar talks, workshops, and conferences. I pursued my fieldwork in Germany, at observatories in Spain and Chile, at a research institute in the United States and at a university in Canada, as well as at conferences, meetings, and at a hack week in Italy, the United Kingdom, the United States, Canada, Germany, France, and Malaysia.

My aim was to witness how astronomers make and use large datasets. They typically do their research in projects, that is, in sequences of work in pursuit of a specific research goal, typically involving several people, extending over months and years, entailing many moments of retrospection and prospection, ideally resulting in publications jointly authored by participating scientists (but, usually, not data-producing technicians). One focus of my ethnography was to follow three PhD students through their thesis projects. Partly overlapping, the other was to witness various stages of the "data life cycle" in action, from the planning and making of observations to their calibration, analysis, and publication. I took detailed fieldnotes and many photographs, and made 434 audio recordings, including 151 interviews, 85 naturally occurring work scenes, 56 work sessions of collaborations, 10 teleconferences, and 52 recordings of academic presentations and discussions.

Even though most of what I witnessed occurred at a small number of sites, *How Data Need People* is not a laboratory ethnography. Much of what I witnessed happened "in-between" places: data were made for users elsewhere; scientists, distributed globally, met in teleconferences, videoconferences, and meetings, and engaged in work done at many places; and papers were written (in part) for unknown readers. Thereby, "the distant" was always present in work that was ostensibly local and the distinction between locality and distance became both problematic and a lens into the workings of data-rich research (cf. Chapter 4).

As I focus on the sequential unfolding of projects over days, months, and years, I have ordered and analyzed my ethnographic material as "streams" that map the temporal, prospective, and retrospective organization of work. In doing so my study differs from those that are based on historical material (Edwards 2010), on surveys (Tenopir et al. 2015), or on interviews (Kriesberg et al. 2013; Leonelli 2016). These are commonly analyzed with approaches like Grounded Theory (Corbin and Strauss 2014), which

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 12 Nov 2026 at 16:31:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.002

cannot attend in detail to the always lived and sequential work of scientists and technicians that interests me. Events like Owen's pun (Transcript I.1) and their interpretation would likely be missing from an interview-based account.

I was attracted to the MAMBO team because of its influential technical innovations. The team analyzed images with semi-automatic software pipelines, made and combined large datasets from telescopes on the ground and in space, developed probabilistic data models, and automated object classification techniques. A few years earlier, team members had drastically improved a technique for estimating photometric redshifts (a measure of cosmic distance) from multicolor photographs (Chapter 3). It enabled them to assemble samples of distant galaxies ten times larger than any other team had produced, and thus to do novel statistical studies of galaxy evolution. When I began my fieldwork, a report on the status of observational cosmology had recently considered it the best-performing survey of its kind worldwide. MAMBO was the foundation for a series of international collaborations, including MUWAGS. Even though the team continued to extend its observing program, it could not compete with a team from the United States that had implemented its technique, had access to an observatory with better weather conditions, and then arguably scooped its key science results. Revisiting the team after 2010, I witnessed its decline. It dissolved in 2016.

One of MAMBO's successes is its afterlife in contemporary science. Its improved photometric redshift technique proved to be an important step for ongoing and planned projects, including analyses of data from the James Webb Space Telescope (as of 2022), the European Space Agency's Euclid spacecraft (launched in 2023), and the Legacy Survey of Space and Time at Vera Rubin Observatory (operating since 2025). The years of my study also mark the widespread use of machine-learning techniques in astronomy. Here again, the MAMBO team's work was prescient as it incorporated methods of probabilistic inference from the start.

My access to the team was facilitated by earlier contacts with some of its members when I was an editor and staff writer of *Spektrum der Wissenschaft*, the German-language edition of *Scientific American*, before I returned to academia. I benefitted from my own training and research experience in extragalactic astronomy, as a MSc student in the 1990s (Hoeppe et al. 1994), before I began to study social anthropology. By doing so I arguably acquired "unique adequacy" to do the present study, the "vulgar competence" that Garfinkel and Wieder (1992) deem necessary for ethnomethodological analysts of work.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 12 Nov 2026 at 16:31:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.002

While my presence as an ethnographer was marked by some astronomers initially ("Now we all have to wear a tie!"), I was quickly taken for granted as a temporary presence and not deemed as someone in need of particular care. My competence and sense of membership became audible to myself, for example, when listening to the recording of the conversation represented in Transcript I.1, in which I heard myself participating in the laughter following Owen's pun. I shared much of the cultural background of the MAMBO and MUWAGS team members and was of a similar age as its postdoctoral scholars and junior faculty members. At times I was a proper participant-observer, such as when I participated in classifying galaxy images for the MUWAGS collaboration, but in many situations of instruction and technical deliberation I tried to be more like a "fly on the wall."

Anthropologists and sociologists debate how the presence of an ethnographer affects a social setting and whether one could then still talk of "naturally occurring situations." I share this concern but note that people need to get their work done – work that other members must eventually recognize as adequate. Witnessing both this work and settings in which others scrutinized it assured me that what I had understood was meaningful to members. Despite the ethnographer onlooking, there is only so much that people can do differently if they want to produce work that their peers will accept.

A further point is how ethnographers can claim any deeper significance for apparently singular events like Owen's pun in Transcript I.1. To generalize from it or claim its representativeness, would one not need to witness, say, five, or ten, or fifteen similar scenes first? This demand for statistics and quantification as sole measures of evidentiality frustrates interaction-minded ethnographers (Crabtree et al. 2013). As Wes Sharrock and Bob Anderson (1986, 93) point out, "from ethnomethodology's point of view, the location of regularity has much more to do with demeanour than with statistical representativeness," and they elaborate: "We do not determine that something is commonplace because we have witnessed thousands of occurrences of it. Seeing it once might be enough to establish for us just how commonplace it is" (Sharrock and Anderson 1986, 93). As with Transcript I.1, and as elaborated earlier, what matters is the "inside view" of how participants in a situation react to certain actions and utterances, signaling unwittingly, for example, that a certain action or its absence is commonplace in a given setting. The regularity and generalizability of social actions is not reducible to extraneous criteria like sample size or the duration of fieldwork.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 12 Nov 2026 at 16:31:08, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.002