CHAPTER 8

Encoding Knowledge

Can Data Be Made to Speak for Themselves?

8.1 This Chapter's Plan

How can scientists pass on what they have learned not only *from*, but *about* their data? This chapter proposes to answer this question by considering data as a form of writing and examining its social uses. I argue that scientific datasets do not merely represent information, but can be structured and presented to have a pragmatic function oriented to enable users' understanding. I demonstrate this by describing how the Multiwavelength Galaxy Survey (MUWAGS; a pseudonym) collaboration, discussed in Chapters 6 and 7, designed its catalog – a table of the measured and estimated properties of galaxies –, to guide users to self-correct wrong uses and delimit being held accountable for misuses. Some astronomers argue that catalogs encode their makers' collective knowledge of their data. I examine this claim based on what I witnessed ethnographically, suggest elements of a pragmatics of data reuse, and end with a reflection on socio-computational orders, that is, entanglements of the social and the computational in data-rich science.[1]

As in Chapter 7, witnessing a collaboration's "inner dialogues" allows us to explore otherwise unarticulated assumptions about how data makers and users think and act as practical methodologists. This is another example of taking a problem of data-rich science – here that of making a user-friendly data release while maintaining a balance of securing credit and demonstrating accountability –, considering its "staffing" with people, and following its management ethnographically.

[1] Here I refer to linguistic pragmatics, which is "centrally concerned with how speakers and hearers achieve understanding in and through language" (Koschmann 2011, 435). I informally seek to transpose its program to address scientists' communication with measurements that are typically numerical. Note the distinction between linguistic pragmatics and philosophical pragmatism (e.g., Misak 2016).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 09:41:00, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.010

8.2 Makers and Users, Writers and Readers

A frequent observation in studies of the reuse of scientific data is that aspiring reusers commonly miss what they consider necessary information about the data, putting its uses at stake. Conversely, data makers worry about the misuse of their data.[2] Finding existing data and making them usable can be challenging, since data reusers often need to know more than metadata ("data describing data") provide. Making the diverse contextual information that reusers desire explicit can overwhelm data producers.[3] Paul 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, a process 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, being coproduced by the interactants themselves (Clark 1996). In this view, data need someone to speak for them.

"If only data could speak for themselves," one feels tempted to respond. Putting it this way echoes Socrates' despair about writing, then (ca. 420 BCE) relatively new as a communication medium in Athenian society. In a famous passage of the dialogue *Phaedrus* (Plato 2005), Socrates compares reading a piece of writing with looking at a painting. Much like paintings are unable to respond to questions viewers ask, Socrates finds written words to "point to just one thing, the same each time" (Plato 2005, 63, 275d5). Writing, like a painting, "trundles about everywhere in the same way" and "does not know how to address those it should address and not those it should not" (276e1). These, of course, are concerns of a philosopher whose favorite mode of instruction is spoken dialogue.

One can define 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*" (Leonelli 2016, 77; emphasis in original), but when conceived as digital photographs ("arrays of numbers") or machine-generated "inscriptions" (Latour and Woolgar 1986), digital data may appear to be text-like, as forms of writing.[4] This would make

[2] Zimmerman (2008), Brewer (2017), Gregory et al. (2019), and Pasquetto et al. (2024). Note, however, that data reuses in astronomy are often successful (cf. Plant and Hanisch 2020).

[3] Compare Faniel et al. (2019) and Yakel et al. (2019). On conflicting definitions of data reuse, see van de Sandt et al. (2019).

[4] Bokulich and Parker (2021, 6) notice that, as per Leonelli's definition, "a mice colony could constitute data, if it is taken as potential evidence for claims about a link between genes and behaviour." They prefer to define data as records that are separate from the system or phenomenon under study.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 09:41:00, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.010

data users the readers of texts who may well experience Socrates' despair. Media scholar John Durham Peters (1999) argued that the transmission of writing is fundamentally distinct from dialogical exchanges in copresence. Philosopher Sybille Krämer (2015, 23) concurs when she writes that "[t]ransmission is precisely not dialogical: the goal of technical communication is emission or dissemination, not dialogue. We can thus clearly distinguish between the personal principle of understanding and the postal principle of transmission." In written discourse there appears to be a "lack of recipient accountability" (Deppermann 2015, 61).

The contrast between transmission and dialogue is stark,[5] but various studies suggest a congruence. Mikhail Bakhtin (1986, 165) muses that "the listener (reader, viewer)" is included "in the system (structure) of the [literary] work." Umberto Eco (1979, 7) argues that, to "make his text communicative, the author has to assume that the ensemble of codes he relies upon is the same as that shared by his possible reader." Wolfgang Iser (1980; 1989) notices that the meaning of texts (and not only literary ones) is found not simply in the words, but in the interaction between text and reader. Ethnomethodological studies that regard readers as hermeneutical practitioners who use documentary methods of interpretation support this view (McHoul 1982; Livingston 1995). Historical studies of science supplement this picture, finding that scientists dominantly used letters well into the eighteenth century to communicate data and results (Daston 1991). These letters were often "simulations of conversations" (Bohn 1999). Jumping to the twenty-first century and moving beyond science, social media users routinely bridge the chasm between text and talk-in-interaction that Krämer identifies. Thus, Patricia Bou-Franch and coauthors (2012) identify interactional coherence, a notion of linguistic pragmatics, in commentaries on YouTube videos, whereas David Giles and coauthors (2015, 2017) use conversation analysis to identify sequential structures in blog postings, chatrooms, and YouTube commentaries. William Housley and coauthors (2017) adopts Erving Goffman's (1983) notion of the interaction order in interpreting sequences of messages on Twitter.

More relevant for a study of data reuse in science is Dorothy Smith's (2001, 175–176) suggestion to conceive of the social, organizational, and institutional uses of texts, especially of printed materials, as

By following Hacking (1992a, 48) and considering only the products of data generators as data, my view is yet narrower and more specific.

[5] Arguably, it is even meaningless, since it sets up a false contrast. Thanks to Grey Gundaker for this alert. I describe Peters' (1999) and Krämer's (2015) arguments for illustrative purposes.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 09:41:00, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.010

> text–reader conversations in which, unlike real-life conversations, one side of the conversation is fixed and unresponsive to the other's responses. (…) However the reader takes it up, the text remains as a constant point of reference against which any particular interpretation can be checked. It is the constancy of the text that provides for the standardization effect. (…) Text–reader conversations are embedded in and organize local settings of work. (…) In standardizing one "party" to every text–reader conversation, the terms of all conversations with the "same" text are standardized. Among participants, an open-ended chain is created: text–reader–reader–reader–.

Smith explores the consequences of the spread of "identical copies" to multiple sites. This resembles Bruno Latour's (1986) account of "immutable mobiles." But Smith goes beyond Latour by focusing also on the sequentiality of "text–reader conversations" and by examining organizational and institutional uses, which provide contexts of (social) accountability.[6] This, after all, is a key point in scientific data reuse: that readers are users. What they do with others' data will itself be open for assessment.

As an elementary form of scientific data, measurements are generated procedurally and are, as such, resources for the achievement of intersubjectivity. Data makers and users – as members of a discipline – commonly agree, in principle, on how those measurements are to be made.[7] Remember how, in measuring the luminosities and redshifts of a sample of distant galaxies, specific contexts of accountability mattered to the fixation of Nadine's dataset and how an invisible public of absent evaluators seemed to be curiously present in Nadine's and Otfried's conversations (Chapters 3 and 4). In turn, Nadine was guided to present her measurements in a "statistical language," a "technology of intersubjectivity" (Hacking 1992b, 152).

Scientists typically supplement public data releases by journal articles that describe the data production, processing, and analysis. These papers are meant to instruct users. But can such instructions ever be complete and will data users follow them? After all, studies of technology use have

[6] Harper (1998). Smith argues, furthermore, that there is a movement between what is "locally historic" (Garfinkel et al. 1981) – that is, sequential, reflexive, possibly interactional, and intersubjective – and textually mediated discourse. When "out of action, [textual materials] exist in potentia but their potentiating is in time and in action, whether in ongoing text–reader conversations or in how the 'having read' enters into the organization of what is to come" (Smith 2001, 174).

[7] For example, measuring the position of a star in a pixel image requires finding its center coordinates, whereas measuring its brightness requires adding the amount of its light recorded in several pixels. Both operations follow discipline-specific methods and protocols, including calibrations. They involve assessments of what is "good enough" to complete a measurement and to share it meaningfully.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 09:41:00, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.010

provided ample demonstrations that many users do not consult manuals when setting out to operate new devices or turn to them only at last resort.[8] Such attitudes have inspired designers to develop artifacts that aspire to be "self-explanatory," that is, "their operation should be discoverable without extensive training, from information provided on or through the machine itself" (Suchman 2007, 43). Yet even when users try to follow such instructions they are bound to be challenged. Revisiting her influential study of how users of a photocopy machine interact with its support system, Lucy Suchman (2007, 4–5) concludes that "human–machine communications take place at a very limited site of interchange," whose asymmetries "profoundly limit possibilities for interactivity, at least in anything like the sense that it proceeds between persons in interaction." Suchman builds on, and illustrates, Harold Garfinkel's (2002) insight that all instructions are essentially incomplete and context-dependent. This finding extends to scientific practice (Lynch and Jordan 1995) and it seems likely that it pertains to data reuses as well.

8.3 Fixating a Catalog, Encoding Collective Knowledge?

Conceived as a form of writing, data can be formatted in diverse ways, from single numbers to lists, tables, matrices in several dimensions, and various relational structures.[9] In astronomy, catalogs are a dominant form of data. At the time of writing this, in August 2025, the Centre des Données astronomiques de Strasbourg (France), a major astronomical data center, provided digital access to 26,503 astronomical catalogs.[10] Many catalogs are tables that list information for one object per row. Columns typically begin with object identifiers (such as the object number) and continue with the celestial coordinates Right Ascension (which resembles geographic longitude) and Declination (which resembles geographic latitude). These are typically followed by columns of specific measurements and their errors, such as the brightness (magnitude) in certain wavelength bands, the shape of objects, radial velocities, photometric redshifts, and so on. Figure 8.1 shows an excerpt of George Abell's (1958) catalog of 2,712 galaxy clusters in the northern and equatorial sky. Abell spent many months visually inspecting the Palomar Observatory Sky Survey's 879 pairs of photographic plates

[8] See Novick and Ward (2006) and Blackler et al. (2016).
[9] For an overview of the variety of data structures, see McKinney (2017) and en.wikipedia.org/wiki/List_of_data_structures (accessed July 7, 2025).
[10] See Genova (2018) and cds.unistra.fr/fr/ (accessed July 7, 2025).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 09:41:00, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.010

No.	R.A. (1855)	Decl.	Precession R.A. (1900)	Decl.	l	b	Mag.	Dist.	Rich.
1	00 00.1	15 43	0.512	3.34	75.6	-45.4	17.1	5	1
2	00 01.0	-20 27	0.512	3.34	35.9	-78.2	17.3	6	1
3	00 01.8	03 14	0.512	3.34	71.1	-57.6	17.0	5	1
4	00 01.9	05 59	0.512	3.34	72.6	-54.9	17.8	6	1
5*	00 02.8	32 18	0.514	3.34	80.6	-29.2	17.1	5	1
6	00 03.0	16 54	0.513	3.34	77.0	-44.4	17.5	6	2
7	00 04.2	31 37	0.515	3.34	80.8	-30.0	17.1	5	1
8	00 04.7	-12 00	0.511	3.34	59.3	-72.0	17.2	5	1
9	00 04.9	08 40	0.513	3.34	74.9	-52.5	18.0	6	1
10	00 05.4	-06 48	0.511	3.34	65.9	-67.3	17.2	5	2
11	00 05.4	-17 15	0.511	3.34	49.5	-76.6	17.2	5	2
12	00 05.9	-08 25	0.511	3.34	64.6	-68.9	17.2	5	2

Figure 8.1 The first entries of George Abell's (1958) catalog of 2,712 galaxy clusters in the northern and equatorial sky, based on his visual inspection of the photographic plates of the Palomar Observatory Sky Survey. For each object the columns list the catalog number (column 1), celestial coordinates and positional information (columns 2 to 7), the visually estimated magnitude of the tenth brightest cluster galaxy (columns 8), as well as coarse estimates of the cluster distance (column 9), and of the number of galaxies it contains (column 10). (© American Astronomical Society. Reproduced with permission)

of the northern and equatorial sky.[11] His catalog was printed on forty-three pages in an issue of the *Astrophysical Journal's Supplement Series*.

If Abell's catalog is remarkable for having been made by a single astronomer, then the catalogs of the Sloan Digital Sky Survey (SDSS) are noteworthy for being the joint work of a large team. The SDSS Early Data Release (EDR), which records measurements of 14 million detected objects, was accompanied by a data release paper published by 192 authors (Stoughton et al. 2002). Thirty-two of these can be identified as lead authors, the core team of catalog makers.[12] (The SDSS catalogs cannot be shown as the Abell catalog, as they are accessed by database queries specific to user interests; Figure 8.2 illustrates instead the formatting of a recent, less complex catalog.) SDSS users can access the survey's "raw" data, such as digital photographic exposures, and process these data "from scratch" according to the specific requirements of their research project. Indeed, doing so is what David W. Hogg, a member of the SDSS collaboration, recommends, in principle, to users who want to exploit the information content of the

[11] Each field "was photographed twice, once on a [blue-sensitive Kodak] Eastman type O emulsion and once on a type E emulsion through a red Plexiglass filter" (Abell 1958, 121).

[12] The author list of Stoughton et al. (2002) is strictly alphabetical only from the thirty-third listed author, suggesting that the first thirty-two authors were particularly involved in the catalog's creation.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 09:41:00, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.010

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 09:41:00, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.010

2MRS Catalog

(1) 2MASS ID	(2) R.A.	(3) Dec.	(4) l	(5) b	(6) K_s^0	(7) H^0	(8) J^0	(9) $K_{s,t}^0$	(10) H_t^0	(11) J_t^0	(12) $\sigma(K_s^0)$	(13) $\sigma(H^0)$	(14) $\sigma(J^0)$	(15) $\sigma(K_{s,t}^0)$	(16) $\sigma(H_t^0)$	(17) $\sigma(J_t^0)$	(18) $E_{B\!V}$	(19) r_{iso}	(20) r_{ext}	(21) b/a	(22) flags	(23) Type	(24) t_src	(25) v	(26) $\sigma(v)$	(27) cat	(28) Bibcode	(29) Catalog ID
	(deg)		(deg)		(mag)								(mag)				(mag)	($\log_{10}$")						(km/s)				
00424433 + 4116074	10.68471	41.26875	121.17430	−21.57319	0.797	0.929	1.552	0.743	0.881	1.497	0.016	0.016	0.015	0.017	0.017	0.016	0.683	3.208	3.491	0.473	Z111	3A2s	ZC	−300	4	N	1991RC3.9.C...0000d	MESSIER 031
00473313 − 2517196	11.88806	−25.28880	97.36301	−87.96452	3.815	4.132	4.858	3.765	4.077	4.798	0.016	0.015	0.015	0.017	0.016	0.016	0.019	2.799	2.965	0.264	Z111	5X s	ZC	243	2	N	2004AJ ....128...16K	NGC 0253
09553318 + 6903549	148.88826	69.06526	142.09190	40.90022	3.898	4.131	4.784	3.803	4.043	4.690	0.016	0.016	0.015	0.018	0.018	0.016	0.080	2.688	2.878	0.517	Z111	2A2s	ZC	−34	4	N	1991RC3.9.C...0000d	MESSIER 081
13252775 − 4301073	201.36565	−43.01871	309.51639	19.41761	3.948	4.244	4.931	3.901	4.203	4.876	0.015	0.016	0.015	0.016	0.017	0.016	0.115	2.445	2.613	0.957	Z111	−2 P	ZC	547	5	N	1978PASP...90..237G	NGC 5128
13052727 − 4928044	196.36366	−49.46790	305.27151	13.34017	4.471	4.790	5.508	4.421	4.735	5.444	0.016	0.016	0.015	0.017	0.017	0.016	0.176	2.627	2.772	0.308	Z111	6B s	ZC	563	3	N	2004AJ....128...16K	NGC 4945
01335090 + 3039357	23.46210	30.65994	133.61024	−31.33081	4.477	4.697	5.346	4.087	4.329	5.003	0.020	0.018	0.017	0.044	0.038	0.029	0.041	2.699	3.032	0.792	Z111	5A4s	ZC	−179	3	N	1991RC3.9.C...0000d	MESSIER 033
09555243 + 6940469	148.96846	69.67970	141.40953	40.56710	4.636	5.003	5.744	4.610	4.973	5.704	0.015	0.015	0.015	0.015	0.015	0.015	0.156	2.357	2.542	0.396	Z111	0	ZC	203	4	N	1991RC3.9.C...0000d	MESSIER 082
03464851 + 6805459	56.70214	68.09611	138.17259	10.57999	4.682	4.952	5.494	4.362	4.682	5.169	0.020	0.019	0.018	0.043	0.040	0.033	0.558	2.571	2.876	0.858	Z111	6X2T	ZC	31	3	N	1999PASP..111..438F	IC 0342
13370091 − 2951567	204.25383	−29.86576	314.58353	31.97269	4.721	4.951	5.594	4.595	4.832	5.480	0.017	0.016	0.016	0.025	0.019	0.018	0.067	2.495	2.709	0.825	Z111	5X2s	ZC	513	2	N	2004AJ....128...16K	MESSIER 083
12395949 − 1137230	189.99789	−11.62307	298.46094	51.14923	4.991	5.228	5.897	4.944	5.177	5.841	0.015	0.015	0.015	0.017	0.015	0.015	0.051	2.305	2.473	0.682	Z111	1A P	ZC	1024	5	N	2000MNRAS.313..469S	MESSIER 104
00424182 + 4051546	10.67427	40.86517	121.14999	−21.97622	5.084	5.301	6.171	5.040	5.275	6.142	0.015	0.015	0.015	0.017	0.016	0.016	0.155	2.168	2.360	0.913	Z111	−6	ZC	−200	6	N	2000UZC...C.....0F	MESSIER 032
12505314 + 4107125	192.72145	41.12015	123.36211	76.00777	5.163	5.408	6.068	5.100	5.344	6.010	0.015	0.015	0.015	0.016	0.016	0.015	0.018	2.236	2.414	0.847	Z111	2A3R	ZC	308	1	N	1993A&A...272...63M	MESSIER 094
12564369 + 2140575	194.18207	21.68266	315.68127	84.42287	5.381	5.623	6.300	5.315	5.558	6.231	0.016	0.015	0.015	0.017	0.016	0.016	0.041	2.332	2.490	0.583	Z111	2A T	ZC	408	4	N	1991RC3.9.C...0000d	MESSIER 064
20345233 + 6009132	308.71805	60.15368	95.71873	11.67289	5.424	5.921	6.147	5.248	5.711	5.971	0.018	0.017	0.017	0.034	0.029	0.025	0.342	2.402	2.680	0.770	Z111	6X1T	ZC	40	2	N	2008MNRAS.388..500E	NGC 6946
12294679 + 0800014	187.44499	8.00041	286.92224	70.19597	5.498	5.732	6.370	5.388	5.622	6.254	0.017	0.016	0.016	0.025	0.021	0.019	0.022	2.253	2.496	0.913	Z333	−5	ZC	997	7	N	2000MNRAS.313..469S	MESSIER 049
13295269 + 4711429	202.46957	47.19526	104.85159	68.56084	5.589	5.796	6.486	5.484	5.632	6.370	0.017	0.016	0.016	0.025	0.020	0.019	0.035	2.296	2.549	0.902	Z111	4A1P	ZC	463	3	N	1991RC3.9.C...0000d	MESSIER 051a
12185761 + 4718133	184.74008	47.30372	138.31985	68.84251	5.592	5.831	6.498	5.458	5.706	6.361	0.016	0.016	0.016	0.022	0.021	0.018	0.016	2.421	2.662	0.495	Z333	4X s	ZC	448	3	N	1991RC3.9.C...0000d	MESSIER 106
03224178 − 3712295	50.67412	−37.20820	240.16275	−56.68984	5.681	5.947	6.547	5.580	5.860	6.427	0.016	0.015	0.015	0.019	0.018	0.017	0.021	2.220	2.470	0.792	Z133	−2X P	ZC	1760	10	N	1998A&AS..130..267L	NGC 1316
13154932 + 4201454	198.95554	42.02929	105.99706	74.28773	5.722	5.947	6.682	5.602	5.818	6.554	0.016	0.016	0.016	0.021	0.020	0.018	0.018	2.310	2.541	0.660	Z000	4A3T	ZC	484	1	N	2008MNRAS.388..500E	MESSIER 063
02424077 − 0000478	40.66988	−0.01329	172.10397	−51.93358	5.800	6.266	6.985	5.776	6.238	6.937	0.015	0.015	0.015	0.016	0.015	0.015	0.033	1.978	2.162	0.880	Z111	3A T	ZC	1137	3	N	1999ApJS..121..287H	MESSIER 077
12434000 + 1133093	190.91670	11.55261	295.87354	74.31767	5.816	6.064	6.740	5.730	5.984	6.647	0.016	0.016	0.015	0.021	0.017	0.017	0.026	2.166	2.383	0.891	Z111	−5	ZC	1117	6	N	2000AJ....119.1645T	MESSIER 060
03171859 − 4106290	49.32750	−41.10807	247.52402	−57.04243	5.847	6.093	6.731	5.653	5.958	6.512	0.016	0.015	0.015	0.024	0.021	0.019	0.013	2.157	2.513	0.891	Z111	0B s	ZC	788	45	6	20096dF...C....0000J	g0317186-410629
11054859 − 0002092	166.45247	−0.03590	255.53194	52.82921	5.848	6.094	6.781	5.763	5.994	6.686	0.016	0.016	0.015	0.021	0.018	0.017	0.057	2.216	2.432	0.693	Z111	4X3T	ZC	801	3	N	2004AJ....128...16K	NGC 3521
00402207 + 4141070	10.09198	41.68530	120.71631	−21.13871	5.866	6.099	6.662	5.557	5.815	6.374	0.020	0.019	0.017	0.045	0.040	0.029	0.085	2.450	2.760	0.594	Z111	−5	ZC	−241	3	N	1991A&A...246..349B	MESSIER 110
12304942 + 1223279	187.70593	12.39110	283.77777	74.49104	5.896	6.144	6.806	5.804	6.060	6.699	0.016	0.016	0.015	0.019	0.018	0.017	0.023	2.134	2.368	0.990	Z111	−4 P	ZC	1307	7	N	2000MNRAS.313..469S	MESSIER 087

Figure 8.2 Structure of the 2MASS Redshift Survey (2MRS) catalog, which contains data on 44,599 nearby galaxies selected from the catalog of 2MASS, a near-infrared all-sky survey, and is supplemented with spectroscopic observations by John Huchra and his collaborators. For each object the columns list an identity number (column 1), celestial and galactic coordinates (columns 2 to 5), measured magnitudes in six infrared bands and their errors (columns 6 to 17), the galactic reddening (column 18), angular size and orientation (columns 19 and 21), flags (column 22), galaxy type (column 23), redshift and redshift uncertainty (columns 24 and 25), as well as additional information (columns 26 to 29). This is not a regular excerpt of the catalog but a portion shown "for guidance regarding its format and content" (Huchra et al. 2012, 6). (© American Astronomical Society. Reproduced with permission)

SDSS's photographic exposures maximally. But for most users the SDSS catalogs have an inestimable benefit, as Hogg explains:[13]

Transcript 8.1

The most important thing about catalogs is … they encode the collective knowledge of the people who make the data. So the Sloan catalog is the only place ((*where*)) we really encoded what we think the noise model of Sloan is … what we think the point-spread function is … what we think the data artifacts are. Because the catalog has been made sensitive to those things. (…) We shouldn't be passing forward these important metadata through the catalog. But the reality is … we are … this is how we propagate these metadata![14]

Hogg's claim, that catalogs "encode the collective knowledge of the people who make the data," may, in one sense, appear to be self-evident and uncontroversial. Of course, scientists ought to have made the best possible use of their knowledge in processing and analyzing their data. They are most familiar with the detectors and procedures they used. They collaborate in teams to make these data, and divide their labor, so would they not use their collective knowledge for everything they release? Furthermore, many astronomers agree that it is virtually impossible to describe all decisions that led to a final catalog. Peter made this point in Transcript I.2 (Introduction), and Otfried concurs when he explains:

Transcript 8.2

You just cannot document all … individual steps … ehm … so that you could give someone the raw data … and she gets the same catalog in the end. That is … ehm … there are too many steps in between that remain undocumented. It's just like that. And that … you cannot really avoid that because this work would grow unfathomably … so to say … I mean the work of documenting what you have done.

In this sense, the practical circumstances of catalog production challenge the distinction between data and metadata ("data describing data") that

[13] See the Appendix for a complete list of the transcription conventions adopted.

[14] David W. Hogg, "Probabilistic Catalogs and 'Catalog Space,'" Space Telescope Science Institute, Baltimore (Maryland), June 16, 2011. See realserver4.stsci.edu/t/data/2011/06/2642/DavidHogg061611.mp4 at minute 12 (accessed February 5, 2013). Noise models (which quantify the effects of the length of exposure time and detector properties on the image quality), point-spread functions (which quantify the blurring of starlight due to atmospheric turbulence and telescope and camera optics), and image artifacts (present in any astronomical exposure) are specific to the processing of digital exposures, such as those generated by the charged-coupled devices of the SDSS telescope at Apache Point in New Mexico (USA). The SDSS catalog's metadata ("data describing data") only include basic general information on exposure times, column headings, etc. Users find more detailed, but not exhaustive, information on SDSS data processing in the data release papers which accompany every new release.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 09:41:00, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.010

is essential to many discussions of data reuse and "open science" (cf. Mayernik 2019).

But Hogg's remarks, although made informally, seem profound in yet another sense and worthy of further reflection: Would "collective knowledge" – an elusive and controversial topic of philosophical debate – here become conceivable, consequential, and meaningful through its materialization in a digital object? Much of the philosophical debate revolves around whether the notions of "knowledge" and "knowing" can be applied not only to individuals, but also to collectives. Another question is whether "knowledge" should be conceived as a proposition, as most philosophers of science maintain, or as a capability, as those following Ludwig Wittgenstein (2009 [1953]) and Gilbert Ryle (1949) argue. To examine these questions empirically, let us first consider some key episodes in the fixation of the MUWAGS catalog.

8.4 Steps in the Collaborative Fixation of an Astronomical Catalog

We know from Chapters 6 and 7 that MUWAGS was an international team of astronomers that made a multiwavelength dataset of the galaxy supercluster A2713 to investigate its dark matter content and its environmental impact on galaxy evolution. Much of MUWAGS's collaborative work was oriented to making a master catalog: a table of measured and estimated properties of around 88.000 objects detected in optical images of the A2713 field. In its production, three main constituent datasets were to be combined: the MAMBO team's optical ground-based data, the Hubble Space Telescope (HST) Advanced Camera for Surveys (ACS) team's measurements of galaxy shapes, and the infrared team's Spitzer Multi-Band Imaging Photometer (MIPS) observations. The final MUWAGS data release included the master catalog and the processed images, as well as maps of weak gravitational lensing in the A2713 field. The catalog was presented as a FITS (Flexible Image Transport System; see Chapter 1) table and accompanied by a data release paper in a leading journal. The data were made available through the Space Telescope Science Institute's archive, the Centre de données astronomiques de Strasbourg (France), and the collaboration's website.[15]

In the following, I give a chronological account of how the MUWAGS team agreed on the final version of its master catalog for public release. Prior to these episodes, the team had resolved dependencies between

[15] See archive.stsci.edu and cds.unistra.fr/fr/ (accessed July 7, 2025).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 09:41:00, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.010

constituent data subsets, as described in Section 6.6.1. None of the following episodes resulted in a return to data calibrations and analyses. They rather specify some methods of "packaging" data (cf. Leonelli 2016).

The following episodes are particularly rich in specific detail necessary to identify and describe these scientists' methods and accounting practices. Readers may wish to jump to Section 8.5 and return for details thereafter.

8.4.1 *Guiding Catalog Users by Introducing and Structuring Redundancies*

One way to counteract potential misuses of a catalog is to introduce opportunities for instructing users beyond the prescriptive information provided in the data release publication. Introducing redundant catalog entries are one means to do so.

In the work of MUWAGS, a major step in moving toward the first catalog version merging the MAMBO (pseudo-acronym), HST ACS, and Spitzer MIPS data was the assemblage of the so-called J2007d catalog.[16] This table of around 88,000 rows (one for each detected object) contained the outputs of diverse algorithms – including optical positions, radiation fluxes, and redshifts – and their error estimates. It included duplicate information and various cross-checks, many comprehensible only to their makers. With 709 columns per object, this draft catalog was too big to be shared meaningfully with any user beyond the team. At a collaboration meeting the number of columns was to be reduced to about 200 and the catalog was to be made more user-friendly.

Peter, the main author of the MAMBO catalog, argued that users working with the flux measurements and their "galactic reddening" would benefit from a redundancy in the catalog. Astronomers agree that dust and gas in our Milky Way Galaxy scatter the light of distant objects and "redden" it. The amount of reddening depends on where distant objects are on the sky relative to the band of the Milky Way, the plane of our (spiral) galaxy. Objects behind the Milky Way band are reddened most strongly. This galactic foreground reddening can be estimated and subtracted from flux measurements.[17] In astronomical terms, de-reddening "delocalizes"

[16] Catalog version names with an initial J are playful allusions to the epoch of celestial coordinates, such as the commonly used epoch J2000. Celestial coordinates change over time due to shifts in the orientation of the Earth's axis.

[17] Since the MUWAGS field is relatively small (about the size of the full moon), one singular galactic reddening correction could be applied to the measured radiation fluxes of all objects in the field.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 09:41:00, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.010

the data, yielding the radiation fluxes one would measure when, hypothetically, observing extragalactic objects from outside our own galaxy. This makes it easier to compare the fluxes and colors of objects observed at different positions on the sky. Every practicing astronomer should be able to calculate de-reddened and re-reddened fluxes; this is taught in introductory laboratory courses for undergraduate students. Thus, team members could have simply listed the dust reddening correction that they had applied in the data release paper, which is what they did. But here, as elsewhere throughout the discussion, there is the lingering expectation that catalog users will not read the data release paper carefully, be prone to make mistakes, and "pester" the team with inquiries and requests. Consider this exchange between Peter and Mallory, the team's principal investigator:

Transcript 8.3

1 Peter: So … um … I thought … it would be best … to make it … as easy and non-confusing as possible for the public … uhh … by giving them something that is de-reddened … and … uh … well … the calibrations are completely consistently corrected and so on … so it's a dataset where you don't need to understand … any of the intricacies … I thought that would be the best because that way we avoid a lot of discussion … but we were also saying

2 Mallory: You avoid people pestering you

3 (Group): hhhhh *((chuckling))*

4 Peter: Sorry?

5 Mallory: You avoid people pestering you ((*chuckling*))

6 Peter: Exactly! yeah … no … seriously … I mean …

7 (Group): haha hhhh

8 Peter: whenever there's a source of confusion it reduces … uh … the motivation of people to use the data … it it it it causes … a lot of … uh … questions … via email from people who are persistent in … in their attempt to use them and it just creates work for everyone … at a potentially perpetual level … and if we can just preempt all of that … that would be the best … and an and … I'd be happy to do this in in very little … time I believe … um … but I think we still need to give them … some … um … we'll need to give them a prescription … prescription of how to … how to re-redden the data

Peter wants to list de-reddened fluxes to make uses of the catalog "as easy and non-confusing as possible" (line 1). Mallory suggests that by doing so he seeks to preempt being "pestered" by users (lines 2 and 5), which Peter

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 09:41:00, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.010

confirms and formulates as the catalog makers' and users' shared concern (lines 6 and 8). Peter and Mallory worry about catalog users' potential mistakes. Interactions with users seem possible, but, worried about their "potentially perpetual level," Peter rejects them in the pursuit of achieving the project's closure.

After agreeing as a team that the catalog would list de-reddened optical fluxes, Peter suggests also including a column with the uncorrected ("un-de-reddened" or "re-reddened") fluxes in the R band, the deepest image of the optical dataset, even though this information was redundant:

Transcript 8.4

1	Peter:	(…) and … so … we need one magnitude that is somehow in common or is translatable
2	Mallory:	Mm hm
3	Peter:	and … uh … the best thing would … be … to … well … look … oh no … it's actually not a problem because I've just suggested we give them two sets of … no … I said … I … we … I suggested we give them one set of magnitudes de-reddened and a prescription how to … re-redden them … uhm … what if we give them this prescription … but one additional actual column
4	Mallory:	Mm hm
5	Peter:	which is … the re-reddened … R-band magnitude … as observed
6	Mallory:	Yep
7	Peter:	because that is the cut that we use in our … em … I mean the … variable that we … use to define cuts on our own and then people can use that same variable … they can do an R equals 24 ((*magnitude*)) cut or something … and if that means 23.786 in R for de-reddened … then … well … so be it … um … at least they would have a column to straightforwardly use … and … and they could see … for the example of this column as well … if their re-reddening procedures with the reddening law … although it'll be a very simple one … has actually worked … they could confirm with this column … "Yes … I'm not doing something … I'm not doing multiplication instead of division or something stupid" … So that would be an extra column … Mm hm
8	Mallory:	Uh … okay … so … I think that sounds reasonable … my … my strongest motivation is that … we … are internally consistent … I mean … are consistent with
9	Peter:	Right
10	Mallory:	what we already published

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 09:41:00, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.010

By including both the de-reddened and the uncorrected fluxes in the catalog, Peter aims to give catalog users the opportunity to assess their calculations with catalog entries. Peter makes this explicit when (in line 7), using reported direct speech, he takes an imagined user's perspective. Mallory approves of including this extra column (lines 8 and 10), emphasizing her concern for the consistency of the catalog with the team's published work. Including the extra column did not affect this.

This was one of several exchanges in which team members pondered introducing redundancies into the catalog. Eliciting a response generates a sequence of actions. Including the extra column enables, and arguably invites, catalog users to perform a three-part sequence: (1) being instructed to re-redden galaxy magnitudes; (2) using these to calculate specific re-reddened magnitudes; and (3) being afforded the opportunity to self-assess their results for one waveband (the R band). This sequence accommodates users' projectable actions. It is reminiscent of a common feature of instructional sequences in classroom talk: a teacher asking a student a question, the student responding with an answer, followed by the teacher's subsequent assessment of this response. This is the I-R-E (Initiation-Response-Evaluation) sequence, or Question with Known Answer (Mehan 1979).[18] Unlike classroom interaction, of course, catalog makers and users are typically not copresent, users' "response" is in writing, and users would self-evaluate their computations.

As Peter did here, MUWAGS team members commonly adopted the perspective of imagined users in their discussions. Given that these astronomers themselves used other scientists' data, this is hardly surprising. Peter told me:

> *Transcript 8.5*
> Every now and then I use other people's data and want to do science with them … want to write papers … and there are factors that interest me as a catalog user. I have caught myself thinking … "Gosh … now it's getting too complicated with this catalog. What all do I have to know to use it properly and not come up with nonsensical interpretations … biased results?" Perhaps the catalog makers have provided lots of descriptive knowledge or whatever … but for me the situation may become uncertain as I don't know how to use this knowledge and use the catalog to transform it into the product that I wanted to have. And then I sit there and wonder: "Isn't what I am actually looking for there somewhere on the web?" And then I use it and that's it. Or I let this paper go because the effort is getting too big.

[18] See also Macbeth (2003) and Lindwall et al. (2015).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 09:41:00, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.010

Peter describes himself as being an impatient reader of other scientists' catalog descriptions, arguably missing the guidance of the (numerical) catalog entries themselves. A written description alone, it seems, leaves open too many ways of going astray.

8.4.2 Numerical Order as a Resource for Making Nonsensical Data Uses Perspicuous

Let us now turn from anticipated sequential orders of action to uses of numerical order for structuring catalog users' experience. A few months prior to MUWAGS's public data release, when team members had made a comprehensive draft catalog, some of the catalog's entries could not be filled. Some objects were too close to bright stars or the edge of an exposure, challenging precise photometric measurements, and others were affected by cosmic ray hits and other artifacts. But due to formatting requirements, table cells could not remain empty. Thus, blanking values – a sort of placeholder – had to be adopted for them. One MUWAGS sub-team had chosen –99 as a blanking value, a number that, they believed, would not be confused with astronomical measurements. Another sub-team had used NaN ("Not a Number") as the blanking value.[19] Users could search for this text string, such as when aiming to delete questionable table entries from an application. Now striving for a consistent catalog that employed a unique blanking value, Peter, the maker of the ground-based optical catalog, sent an email to the team, requesting ideas for how best to pick a single, consistent blanking value for the entire catalog. His message elicited a lively exchange: over two days, nine team members sent twenty-three messages, from which I quote in what follows.

Ben and Mike, in charge of the optical HST catalog, were mostly concerned with their code's operability and argued for a blanking value of –99 because IDL (Interactive Data Language), popular for data analysis, could not easily process NaNs. By contrast, Susheela and Mallory emphasized that catalog users ought to be able to recognize their mistakes, such as when accidentally calculating with a numerical blanking value like –99. Susheela writes:

> I prefer "NaN" as it limits possibility of error from the user, with absolute mag ((*magnitude*)) cuts and a global replace by the user to his/her blanking

[19] In the Institute of Electrical and Electronics Engineer (IEEE) standard for floating-point arithmetic (IEEE 754), NaN is a number type for undefined or unrepresentable values. See MacKenzie (1993) for a historical sociology of the IEEE 754 standard.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 09:41:00, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.010

> value is clean and unambiguous. I agree with Mike that we should not mix blanking values.

Mallory writes:

> Although I am used to the –99 values I would prefer NaN in this case. Users not realizing the convention can get easily caught out, e.g. with a magnitude selection such as R<24. NaN avoids this so seems better practice to me, and can be easily globally replaced if required.

Susheela and Mallory are both alert to the dangers of confusing blanking values with physically meaningful table entries (in this case astronomical magnitudes). Both argued that NaNs, because of not being numbers, would lead computer code to crash, thereby making wrong uses recognizable. This seemed clearly preferable to introducing unrecognized mistakes. However, Chuang notices that this is not always the case when using NaNs as a blanking value. He realized that some code converts NaNs into numbers that would be difficult to trace, making it challenging to recognize mistakes:

> I have a slight preference for –99 because it won't cause any *algorithmic* difficulties for unwary programmers now and in the future. If you program outside of IDL, MIDAS,[20] or what not, you won't have to worry about it causing your program to crash. I do nearly all my programming in C, for example, and C will convert NaN into a float ((*floating point number*)) or int ((*integer number*)), depending on the conversion string you choose, but it will not otherwise complain that NaN is not a number. What then happens is that NaN becomes some number that gets operated on. It'll likely produce a nonsensical result when that happens. But depending on C or other languages to produce a nonsensical result is quite dangerous, when the sense of the nonsense is not in our control (does this make sense?). So if we're exporting the catalog to the outside world, we ought to keep this in mind.
>
> My feeling is that mathematical operations on –99 would produce numerical nonsense that is in our control and would be easy to flag. Otherwise, choosing an even more unrealistic number, e.g. –1e99, should solve any chance of confusion.

Here –1e99 = -10^{99} is an immensely large negative number, far outside the range of astronomical quantities measured or calculated in this team's work. Its mistaken use would be easily spotted in any calculation.

[20] MIDAS, the Munich Image Data Analysis System, is a software package for astronomical data analyses that the European Southern Observatory provides for researchers; see eso.org/sci/software/esomidas/ (accessed July 7, 2025).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 09:41:00, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.010

Writing independently of Chuang, Otfried came to a similar conclusion and proposed to pick a blanking value "outside the range of all possible data", that is, an "unrealistic number":

> I understand that NaN is not perfect for some packages (especially those, that use ASCII as input) So a VERY negative number seems preferable. But I would strongly argue for a number which is not just outside the magnitude range, but well outside the range of all possible data (for instance distance vectors between two objects in arcsec could be about –1800.) So I suggest something more substantial: –99999 or so.

But even such a large blanking value could be problematic, as Mike points out in response to Chuang and Otfried. Mike argued that globally replacing a table entry would depend on which number formats users would choose. It would produce values that may not be recognized easily. Rounding errors could make it hard to recognize them as well:

> When doing a global replace the maximum allowed data range is specified by the "smallest" data type, in our case integer. So using –99999 will produce –32768 e.g. for NR [the catalog object number]. We should keep that in mind in the choice of our value. Or we would have to consider introducing several values (–9999 for INT [integers], –999999999 for LONG, FLOAT, DOUBLE). Also, keep in mind that for large numbers round-off errors might occur (float(–1e22) = –1.00000e+22; double(–1e22) = -9.9999998e+21).

Mike, who had initially rejected using NaN as a blanking value, now changed his mind, noticing that the popular IDL data analysis code he preferred could, after all, process NaNs:

> On second thought, for IDL users NaN is not that much of a nuisance. If encountering a –99 the plot range has to be adjusted anyway, so filtering has to be done and then one could also use finite() to remove NaNs. NaN also has the advantage that after removal plots can be made without interfering with the plot range at all (min and max has now a sensible value).

Mike's suggestion moved critics like Chuang to accept NaN as the catalog blanking value. Mallory, the principal investigator, subsequently endorsed this with the team's approval. The "README file," a brief users' manual released along with the data, was supplemented to give instructions for how to best find and replace the NaN entries in the MUWAGS catalog.

This discussion illustrates how MUWAGS members were mindful of how imagined users would work with the catalog: doing calculations, checking their results, and noticing what makes astronomical sense and what does not. The team considered, but ultimately rejected, dwelling on a

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 09:41:00, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.010

numerical order that catalog makers and users share as professional astronomers: "the range of all possible data" as Otfried called it.[21] This is a "realist strategy" that distinguishes between what is deemed "real" and what is "artefactual" (Barnes et al. 1996, 81). It provides users with a resource to self-correct their work. As in the previous episode, it is reminiscent of the classroom I-R-E sequence. But in this case the formal constraint of how numbers are represented in the catalog, set by the IEEE 754 standard, thwarted the team's efforts to do so. Any specific numerical blanking value represents a choice for which the MUWAGS team could be blamed. Employing the nonnumerical NaN blanking values arguably shifts the responsibility of using catalog entries properly to users.

8.4.3 Guiding Catalog Users by Selectively Deleting Information and Defining Flags

Besides introducing redundancies to enable catalog users to cross-check their work, catalog makers may delete information that is prone to mistaken uses or mark catalog entries with flags – numerical or textual descriptors that alert users to possible issues with certain entries.

Late in the assembly of the MUWAGS catalog, the infrared-derived galaxy masses and star formation rates, previously listed separately, were merged with the optical MAMBO/HST catalog. Before doing so, Eddie, the head of the Spitzer MIPS infrared sub-team, decided to delete what he regarded as confusing information and applied flags to mark them as not being contained in the survey's area. At a teleconference, his decision became the topic of an exchange between Mallory, Peter, and Eddie:

Transcript 8.6

1 Mallory: Oh … actually … we didn't talk about the star formation rates (hhhh) for MUWAGS … Eddie … you want to say a bit about that … briefly?

2 Eddie: Ah::: … not really. Just to say that they're … they're <u>almost</u> the same as before … you know … so anyone using star formation rates will have to look at it … ahm … I'm assuming that … there'll be <u>some</u> small amount of documentation … if … if not I can send an email … ahhhm … but basically what happened is that there are now total star formation rates for galaxies with infrared detections that are in the <u>high</u> signal-to-noise part of the image …

[21] In Hoeppe (2020b), I examine how some astronomers employed the visual order of digital photographs as a resource for the reuse of data.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 09:41:00, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.010

and I've cleaned up all the low significance detections and all the … detections in … messy parts of the image by simply killing them … ahm … and saying … ahm … and just saying that there is no information in those parts of the images … so …

3 Peter: Can I just ask and confirm? … so when you have a … a low signal-to-noise part of the image or something … ahmm … do you give them flag zero "outside of the area" or do you give them flag two meaning "not detected"?

4 Eddie: I gave them flag zero meaning "outside of the area" and I set

5 Peter: Okay

6 Eddie: I set their … all their infrared-based data to zero

7 Peter: Yeah … okay

8 Eddie: Ahm … so I have gotten rid of it … which is not necessarily the … ahh … the best thing to do for those who are interested in all the data but … I thought that if I did … I mean … I had confused Susheela by doing the other thing … so I figured

9 Peter: Yes. ( ) (complete sample) now

10 Eddie: I choose something simple … ahhh … ahmmm … and should make everyone happy

Invited by Mallory to update the team on his work on the star formation rates, Eddie reports the changes he made to a previous version (line 2). The potentially contentious nature of his action – the "killing" of information derived from "messy" parts of the image – becomes noticeable through Peter's request for explanation (in line 3), which Eddie answers and continues to address (in lines 8 and 10) despite Peter repeatedly acknowledging his understanding and acceptance (in lines 5, 7, and 9). Mallory remains silent throughout this exchange.

When Peter and Eddie talk about "flags" in this exchange, they refer to sample selection and quality flags.[22] Quality flags can be assigned manually to catalog entries or generated automatically by algorithms like the source detection code SExtractor (Bertin and Arnouts 1996) and the GALAPAGOS pipeline (Barden et al. 2012). The MUWAGS collaboration defined quality flags for each of the constituent "sub-catalogs." These flags were refined while fixating the data release and writing the data release paper. At this stage in the discussion the MIPS catalog had three sample selection flags: 0 ("source not covered," that is, not in the MIPS "footprint" on the sky), 1

[22] Table 9 of Stoughton et al. (2002) lists the quality flags adopted by the SDSS's EDR, a model for other survey projects.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 09:41:00, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.010

("source covered and detected"), and 2 ("source covered, but not detected" – that is, a flux density below the detection limit).

As the head of the infrared sub-team, Eddie was entitled to set all entries for "messy parts of the image" (line 2) to zero. What is a "messy" part of an infrared image was not for the members of other sub-teams to judge. However, what Eddie describes as "killing" has a moral connotation. Setting numerical values of the table to zero can be heard as disregarding the epistemic and economic value of these data, obtained as they were using the particularly precious observing time of a space telescope. Eddie acknowledges that deleting entries is not the best thing to do, but he emphasizes his orientation to avoid confusing catalog users, maintaining that even Susheela, a team member, had been confused (in line 8). For Eddie, this concern for cooperation and intelligibility, among team members and beyond, overrides the effort to maximize the catalog's information content. Note that Susheela appears to have made visible to Eddie what he could not presume to be an unquestioned background.

Reflecting on the formulation of quality flags, a collaboration member told me:

> *Transcript 8.7*
> You know … we insert a column for the dumb ones. This sounds arrogant … but what I mean is this … Let's pretend the public is dumb. And what we do is to tell them "Look … this is a column for you … and if you find this number there then just ignore this thing and use the rest only … before it's getting too complicated … where too much can go wrong … where you have to know too much as a user … or where we would have to communicate too much too precisely … and we are not willing to make that effort." … We try to simplify the situation. In that way you cannot use one hundred percent of the power of the catalog … but they can … let me just make up a number … it can be used to eighty or ninety percent by the dumbest possible user. At least nothing will go wrong. That is the point. Better leave opportunities untouched than to let users produce nonsense.

Thus conceived, the resort to flags is a shortcut to account for operations that are difficult to describe and prone to mistaken uses. Eddie's assignment of flags marks the closure of work on the infrared data. These were not processed further.

8.4.4 Turning Makers into Users: Testing the Catalog by Trying to Break It

The end of the team's work on the master catalog was marked by several efforts to "break" it, that is, to check its entries' coherence and logical

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 09:41:00, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.010

consistency and impersonate catalog users beyond the team to identify remaining inconsistencies in need of correction. In Section 6.6.2, I reviewed this work through the lens of organizational practice. After several trials, the team arrived at a catalog that allowed diverse uses coherently, was logically consistent, and was accountable to the various evidential contexts (Pinch 1985) that members investigated. At this point it was "frozen in," as Mallory declared, and became a singular digital object. Given the compromises that fixating the catalog entailed (cf. Episode 3), it was irreducible to any individual team member's work.

8.4.5 Contemplating Additional Possibilities for Users' Sequential Engagement

After the MUWAGS catalog was "frozen in," team members kept pondering further opportunities to engage and instruct users, but the team's discussion also turned to possibilities of holding users to account. At the last collaboration meeting before the data release, team members considered additional means for preempting wrong uses of the catalog. I use quotation marks to transcribe what I heard as quoted fictional text.

Transcript 8.8

1	Mallory:	But I mean we did have discussions on how to … on defensive things … how to guard against lazy and stupid people and possibly making some quantities negative … thinking that that would be very obvious … but it's not necessary. So all we could do is explicitly write down what these quantities are
2	Ken:	Yeah
3	Mallory:	and explain them very well … and point people to everything
4	Ken:	That's a good idea. We can test … you know … fill out a multiple-choice test. "If you want to create … if you want to create … stellar mass versus … something … for cluster members
5	Ben:	Aha!
6	Ken:	which column do you use?"
7	Ben:	The MUWAGS catalog driver's license!
8	(Group):	HA-HA-HA-HA-HA-ha-ha-ha ((*laughter and mumbling*))
9	Mallory:	But that's in light … but it's not necessarily a bad idea … again in the README a few obvious examples … to say … "If you want … what are the obvious things for … stellar masses for confirmed galaxies … or

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 09:41:00, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.010

star formation rates for detected galaxies?" We could actually put that … in as little

10 Elias: and give references to where these things are explained in detail

11 Ken: and say why the other ones are wrong. We could say "Do not use that one because … under your assumptions … don't use that"

12 Elias: Will there be references where this is explained?

13 Mallory: Well … it's all going to be explained in the data release paper. That is going to have everything in there. It's going to be like twenty pages.

14 Elias: Yeah … yeah. It's good as a reference paper. So again … just balance between being comprehensive and then people not reading it because it's too much stuff.

15 Christina: Twenty pages!

16 Mallory: Uh huh. It's not done yet.

In line 1, Mallory alludes to the discussion on blanking values summarized in Episode 2. While she argued for explaining questionable catalog entries in the data release paper, Ken jokingly suggests letting potential users take an online quiz before granting them access to the data (lines 4 and 6). Mallory instead opts for written instructions in the README file, a document that can be downloaded along with the catalog (line 9; cf. Ochsenbein et al. 2000). Much like the "MUWAGS catalog driver's license" (line 7) it would instruct users. In lines 10 and 11, Ken and Elias continue a sentence that Mallory began in line 9, transforming it into a collaborative formulation. Elias expresses the concern, previously made, that users would not read the data release paper with sufficient care (line 14).

This discussion continues the team's playful and exploratory mood of the conversation excerpted in Transcript 7.2 (Chapter 7). There, team members pondered users' signing and acknowledging the "terms and conditions" of the data release as well as requiring them to deposit their credit card numbers for financial compensation in cases of misuse. Somewhat like the introduction of redundant columns in Episode 1, the multiple-choice test that Ken suggests in line 4 is meant to elicit users' responses. It projects users' actions, unavailable for correction in copresence, and seeks to prompt them to self-correct their understanding. References to users passing a driver's license test and committing to a contractual agreement, although made jokingly, can be heard as pointing to team members' desire to hold users legally accountable for their actions.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 09:41:00, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.010

8.5 Discussion

I began this chapter by wondering whether data can be made to "speak for themselves." I examined data as forms of writing that can be "more-than-evidential" when they are structured and presented to have a pragmatic function oriented to users' sensemaking and understanding, with the caveat that writing and written instructions can never substitute for talk-in-interaction. They cannot determine users' actions. Five episodes of the MUWAGS team's "inner dialogues" illustrate how its members aimed to produce an astronomical catalog that was acceptable to all of them, that was coherent with their diverse projects, and that would preempt at least some mistaken uses. In the following, I draw on these episodes to discuss how a data object can instruct potentially unruly users, what they tell us about the claim that a catalog encodes a team's collective knowledge, and how socio-computational orders become a worthwhile topic of ethnographic inquiry.

8.5.1 Making an Instructing Data Object for Potentially Unruly Users

Although it would have been possible for MUWAGS team members to consult with potential users and design their data release accordingly, they did not do so. But real and imagined users (and imagined uses) featured prominently in team discussions, in which members commonly represented users' actual and presumed actions and intentions with reported direct speech (such as in Transcripts 8.4 and 8.5) or referenced them otherwise (such as in the online discussion on blanking values in Episode 2). Imagined users were described as potentially "pestering" catalog makers (Transcript 8.3), as not reading instructions carefully (Transcript 8.8), and as being prone to make mistakes for which they could hold the team accountable (Transcript 8.6). In sum: viewed through their (imagined) actions, users were deemed potentially unruly – and so were (potential) uses of the released data. A closer look reveals that these characterizations draw on team members' self-reflection of their own conduct as professional astronomers. They assume a "reciprocity of perspectives" (cf. Schütz 1962; 1964).

The designers that Steve Woolgar (1991) studied considered users as generic subjects that were to be configured, Madeline Akrich (1992) observed how users were "scripted" into a design, Wes Sharrock and Robert Anderson (1994) witnessed users being typified as "scenic features in design," whereas David Martin and coauthors (2007) witnessed potential users as the "context" for design. But for these astronomers, data makers and anticipated data users were agents who act in recognizably structured

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 09:41:00, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.010

ways as members of the same epistemic community. MUWAGS team members were themselves users of other scientists' data (cf. Transcript 8.5) and they drew on this experience as they assembled their catalog. In a certain sense, team members became (auto-)ethnographers of their own culture, a stance now familiar from Chapters 2, 3, and 7.

Since team members could not draw on resources available for repair in face-to-face interaction, their challenge was to structure the data release to make it an object that guides its users beyond the instructions provided by the data release paper. The episodes described in this chapter reveal some methods for doing so: introducing and structuring redundancies (Episode 1), choosing specific blanking values (Episode 2), selectively deleting table entries and defining flags (Episode 3). Yet other methods could certainly be identified. To speak of methods here may seem exaggerated. After all, little of this was particularly noteworthy for these scientists themselves. Yet it is just their apparent "common sense" that makes these ways of acting a part of extragalactic astronomy's form of life.

Redundancies may, at first, seem odd as a means for communication in science. In a manual for science writers, Silvia M. Rogers (2014, 59) states that "[r]edundancies are common troublemakers in scientific communication." But the diverse uses of redundancies in coding, data storage, cryptography, and communications challenge this view.[23] An instructive case is the use of notation in mathematical writing, such as when formulating exercises. Donald Knuth and coauthors (1987, 19) explain:

> Exercises are some of the most difficult parts of a book to write. Since an exercise has very little context, ambiguity can be especially deadly; a bit of carefully chosen redundancy can be especially important.

Much as students who try to solve textbook problems, data users often lack the experience of using data makers' detectors, algorithms, and analysis procedures. Redundancies are not merely duplicating information but afford users a variety of sequential engagements to assess their understanding. That repetitions of utterances in talk-in-interaction are not meaningless is a central lesson of pragmatic understandings of language. In Episode 1, I noticed team members' aspiration to design redundancies meant to instigate users to go through a three-part sequence for self-correcting mistaken uses. I found this to be reminiscent of student–teacher interaction in classrooms.[24] This triadic structure builds on what Harvey Sacks,

[23] Shannon (1948) and MacKay (2003).
[24] Mehan (1979), Macbeth (2003), and Lindwall et al. (2015).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 09:41:00, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.010

Emanuel Schegloff, and Gail Jefferson (1974, 728–729) refer to as a "proof procedure" for the analysis of turns in conversation, whereby speakers' responses to a previous utterance display their understanding of the utterance to coparticipants of a conversation.[25]

I wondered at the beginning of this chapter if we could formulate elements of a pragmatics of data reuse. Here I understand "pragmatics" inclusively as the study of "meaning in context" (Chapman 2011, 1), and of "how speakers and hearers achieve understanding in and through language" (Koschmann 2011, 435), transposing it to scientists' uses of (numerical) measurements. Sequential engagements with catalog entries, as elicited by means of redundancies, could be one of its topics. Intersubjective uses of measurements, uses of statistical language, statistical descriptors, and the formatting of data could be others. After all, lists and tables afford specific uses and help reducing ambiguity.[26] Their structure, order, and notational characteristics afford diverse uses which are "foundational for coordinating activity distributed in time and space" (Bowker and Star 1999, 138). But, when engaged through accounting practices, other formats afford pragmatic uses as well, including digital pixel images.[27]

8.5.2 *Does a Catalog Encode a Team's "Collective Knowledge" of its Data?*

Inspired by astronomer David W. Hogg's claim that astronomical catalogs "encode the collective knowledge of the people who make the data" (Transcript 8.1), I wondered if "collective knowledge," a topic of philosophical debate, becomes conceivable, consequential, and meaningful (only) through its materialization in a digital object. I now attempt to formulate an anthropological response to this question.

Let me begin by making two distinctions. One is whether "knowing" should be conceived as propositional – that it "consists in possessing the right sort of belief in the right sort of propositions" (Chang 2017, 103) –, which is what most philosophers of science seem to assume, or as a capability, as Wittgenstein (2009 [1953]) and Ryle (1949) argued.[28] The other

[25] See also Moerman and Sacks (1988). Lynch (2011, 554) argues that rather than "offer[ing] a guarantee that a conversation analyst who uses a tape recording to get access to what co-participants' 'understand' from each other's talk will get it right" it only "narrows the field of relevancies" for such analysis.

[26] Goody (1977), Schmidt and Wagner (2004), and Dourish (2017).

[27] See Chapter 4 and Hoeppe (2019b).

[28] Hasok Chang insists that science needs both knowledge-as-ability and knowledge-as-information. Chang quotes Philip Kitcher, who argues that "a crucial feature of propositional knowledge is that

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 09:41:00, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.010

is to distinguish summative views of collective knowledge from those that insist on its irreducibly collective nature. Let me begin with the latter pair, which presumes that knowledge is propositional. The summative view asserts that "a collective knows p iff [if and only if] each member knows p" (de Ridder 2014, 38). Thus conceived, collective knowledge is reducible to the knowledge of individuals and so there is nothing distinctively collective about it. Views that consider collective knowledge as irreducible to that of individuals are of greater interest. For example, a committee could arrive at a certain position without each, or any, individual member subscribing to it (Wray 2017). If one adopts the commonly held philosophical notion of knowledge as "justified true belief" – as all these accounts do or at least set out from – one is led to regard human collectives as "epistemic subjects" that can hold beliefs collectively. This is what Margaret Gilbert (2000) argues for in her "plural subject theory."

Brad Wray (2007; 2017) objects to Gilbert's view by arguing that ascribing shared acceptance to a collective is more plausible than positing a shared belief. He reformulates knowledge as "justified true acceptance." In either case, the collectivity of knowers is delineated by those who believe or accept a claim. Whereas Gilbert extends collective beliefs to disciplines and adherents of Kuhnian paradigms, Wray (2007) confines collective acceptance to research teams and committees, arguing that only these have specifiable decision procedures. Probing another part of the definition of knowledge as "justified true belief," Jeroen de Ridder (2014) suggests attending to the unavoidably collective justification of knowledge in science. It is, for example, beyond any individual scientist's capacity to justify or evaluate findings of elaborate experiments. Questions of collective knowledge then turn into questions of justification and ultimately into questions of trust (Hardwig 1991; cf. also Wagenknecht 2016, ch. 8).

These considerations can inform an interpretation of the MUWAGS catalog's fixation. One may accept both Wray's (2007; 2017) replacement of "belief" by "acceptance" and de Ridder's (2014) unavoidably collective justification of its contents. The data release paper's collective authorship signals the collective acceptance of the dataset that it describes. My five episodes illustrate how the justification of catalog entries and quality flags relied on members of distinct sub-teams. That the final catalog became irreducible to individuals' knowledge is suggested by the diverse evidential contexts engaged by team members as well as the compromises made in its completion.

it is the most suitable form of knowledge for public communication" (Chang 2022, 17). As this chapter suggests, this may not hold for scientists communicating knowledge through data.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 09:41:00, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.010

However, for the naturalistic position that I adopt, the propositional view of knowledge as a form of belief is disappointing. After all, scientists' "[s]hared beliefs are much less common than shared practices" (Netz 1999, 2).[29] It is more interesting to follow Gilbert Ryle (1949, 133–134) and consider "know" as a "capacity verb (…) that is used for signifying that the person described can bring things off, or get things right." By contrast, for Ryle, "believe" is a "tendency verb and one which does not connote that anything is brought off or got right." Ascribing knowledge to someone therefore presumes the witnessability of this person's actions, or of these actions' products. Harold Garfinkel turned Ryle's and Wittgenstein's (2009 [1953]) problems of mutual understanding into topics for empirical study. Responding to fellow sociologist Aaron Cicourel in a discussion on researching children's acquisition of language skills, Garfinkel (in Hill and Crittenden 1968, 47) put it pithily:

> "know" here has to do not with what one might have in mind in some secret place. It is not a case of your having to calm a respondent or seduce him in order for him really to tell you. Then you would be illuminated on what he had been hiding all along. Instead, "know" consists really in a structure of activity. That is what the "know" consists of.

Note that, in holding this view, Garfinkel explicitly counters arguments that locate knowledge in the mind or in the brain – arguably the root cause of the philosophical controversy on collective knowledge. Garfinkel (1967, 30–31) insists that the "appropriate image of a common understanding is therefore an operation rather than a common intersection of overlapping sets." He is concerned with "a *procedural* sense of common or shared, a set of practices by which actions and stances could be predicated on and displayed as oriented to 'knowledge held in common' – knowledge that might thereby be reconfirmed, modified, and expanded" (Schegloff 1991, 152; emphasis in original).

In the MUWAGS team's catalog making, only their shared data analysis skills and practices of agreement may be collective in this sense, since otherwise its members' skills (capacities) were clearly differentiated and irreducible, as the five episodes above and the two episodes in Chapter 6 have shown. It is then the catalog's singularity, its unity, and "thingness"

[29] See also Rouse (2003) and Chang (2017; 2022). Otfried, Nadine's supervisor (Chapter 3), repeatedly expressed to me his belief that the standard Lambda cold dark matter (ΛCDM) cosmology was wrong. The Hubble constant, he believed, must be much lower than the commonly accepted value. Nevertheless, he used the latter in all of his publications. Using a different value may not have passed peer review at important journals and reduced his readership.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 09:41:00, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.010

in its published, materialized, fixated form, that is most interesting for a discussion of "collective knowledge."[30] In this "frozen" form, the catalog is a "means of forgetting" (de Certeau 1984, 97), leaving no trace of the discussions and agreements involved in its closure, a product of its makers' analyses and the team's negotiations. It may properly instruct only users familiar with the evidential contexts, to address which the catalog was designed for, and may be contingent, as well, on shared accounting practices. In its users' data analyses, the catalog makes the team's knowledge of its data actionable, but it does so without developing users' capacity to "bring things off, or get things right" (Ryle 1949, 133) beyond specific applications.[31]

8.5.3 Socio-Computational Orders

Learning from data is the point of making and using them. But how to pass on what has been learned not only *from* data, but about *them*? A "frozen" catalog has "frozen" the knowledge of its makers about their data without specifying it.[32] This, after all, is how its fixation is a "means of forgetting." One way to learn is to change one's beliefs, but "belief" need not be something hidden in a private mind (cf. Ryle 1949). Computational scientists rather use it as in Bayesian inference, where it stands for the prior distribution as an estimate of the probability of a hypothesis. According to Bayes' theorem, a prior is updated by multiplying it with the likelihood – the "measure of the evidential support provided by data for particular parameter values" (Spiegelhalter 2019, 392) –, resulting in the posterior probability distribution. "In a sense," argues Sivia (2006, 6), "Bayes' theorem encapsulates the process of learning." It is a principle foundational not only to the spectral template-fitting technique that Nadine used (Chapter 3), but also to machine learning and generative artificial intelligence.

[30] The philosophical literature has little to say on how the materiality of writing shapes conceptions of collective knowledge, an exception being when Wray (2017) discusses Beatty's (2006) example of how an expert committee arrived at an assessment of radiation safety, which prominently involved the use of documents. By contrast, discussions of collective authorship commonly focus on documents – the authored publications (e.g., Hardwig 1991; Galison 2003; Huebner et al. 2017).

[31] See Ribes and Bowker (2009) for an alternative view of encoding the knowledge of communities, focusing on communally shared and explicable classifications and descriptions (ontologies). Another view is to conceive of a collaboration as the group to which knowledge about its data properly belongs (cf. Sharrock 1974).

[32] See Hogg and Lang (2008, 2010) for intriguing ideas on how to transmit makers' knowledge of their data not through catalogs, but via imaging data and computer code. However, at least in 2025, catalogs were still widely used in astronomy. Transmitting knowledge through code also hinges on the maintenance of the code's usability, which can be challenging in the long run (Cohn 2019).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 09:41:00, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.010

Updating beliefs with Bayes' theorem can appear to be a purely mechanical operation, but it is also deeply social. Not only are Bayesian priors unavoidably subjective, but every inquiry must begin somewhere and where to begin is a human choice. Social agreements, shared assumptions, and normative considerations are the ground on which data are usable and meaningful. Measuring, and agreeing on the adequacy of measurements, is about sharing "common ground."[33] Thomas Kuhn (1970) argued that scientific disciplines are "language communities" bound by "concerted agreements on theories, measuring techniques, and characteristic modes of demonstration" (Lynch 1991b, 105fn 1). As we have seen in Chapter 3, learning to make proper measurements is a key element of achieving membership in such a community. Their social and epistemic orders are aligned, and so is the numerical order of the measurements central to their work (cf. Section 8.4.2). Measurements are presented in a "statistical language" that serves as a "technology of intersubjectivity" (Hacking 1992b, 152).

The collaborative production of a catalog (and other forms of processed data) may encode a team's knowledge about noise models and artifacts, but it also serves future users as a shared reference. It brings them "on a page" and allows them to compare their analyses and assess these analyses' robustness.[34] In each of the SDSS's five-year phases, subsequent annual data releases were meant improve earlier ones, but all data releases are still available online. As Alexander Szalay, a long-term SDSS collaboration member, put it: "A data release is like a book. You cannot take it down. People use old editions."[35] This adds another view to the uses of writing in research with large datasets that this chapter has explored.

[33] Lewis (1969), Clark (1996), and Enfield (2013).

[34] Key elements of many astronomical catalogs are produced with computer code that is widely shared, including the SExtractor source identification code (Bertin and Arnouts 1996) that Nadine used (cf. Chapter 3). In recent years, code intercomparison projects, in which the output of different codes in processing the same dataset is compared, have aimed at agreeing on best practices (e.g., Kitching et al. 2013 and references therein). See also Sundberg (2011) and Mayernik (2021).

[35] Heidelberg Astronomical Colloquium, Heidelberg (Germany), May 22, 2007. Note that subsequent phases of the SDSS do not simply improve on earlier observations but use new instrumentation and observe different targets.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 09 Nov 2026 at 09:41:00, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009686754.010