

# Transforming Taxonomic Interfaces: “arm’s length” cooperative work and the maintenance of a long-lived classification system

ANDREA K. THOMER, School of Information, University of Michigan, USA

MICHAEL B. TWIDALE, School of Information Sciences, University of Illinois at Urbana-Champaign, USA

MATTHEW J. YODER, Illinois Natural History Survey, University of Illinois at Urbana-Champaign, USA

If scientific research can be described as “arm’s length” cooperative work, then the sciences engaged in the creation and maintenance of cooperatively built classification systems might be said to require extremely long arms; workers must reach across great distances of time, space and awareness to independently contribute to the creation and maintenance of a conceptual infrastructure. This is particularly the case for the field of biological taxonomy, in which researchers across the globe have, for more than three hundred years, cooperatively developed a massive knowledge organization system describing all life on earth. As taxonomists move from paper-based workflows to born-digital, data-first modes of publishing, new tools and approaches are needed to support their work. In this paper, we present research conducted as part of the “Transforming Taxonomic Interfaces” project aimed at improving interfaces for taxonomic software. We describe biological taxonomy in the context of CSCW, and identify key strategies taxonomists deploy to facilitate loosely coupled cooperative work. We also contribute a discussion of *semantic refactoring*, a unique kind of “articulation work” entailed in developing, maintaining, and migrating classification systems. This work has implications for the design and study of knowledge organization systems and infrastructure, classification research, and for the development of general semantic web tools and software as well as those specifically for biological taxonomy.

CCS Concepts: • **Information systems** → **Collaborative and social computing systems and tools**; • **Human-centered computing** → **Computer supported cooperative work**; *Empirical studies in collaborative and social computing*; • **Applied computing** → **Life and medical sciences**;

Additional Key Words and Phrases: classification; taxonomy; biodiversity informatics; articulation work; semantic refactoring; “arm’s length” cooperative work; loosely coupled cooperative work

## ACM Reference Format:

Andrea K. Thomer, Michael B. Twidale, and Matthew J. Yoder. 2018. Transforming Taxonomic Interfaces: “arm’s length” cooperative work and the maintenance of a long-lived classification system. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 173 (November 2018), 23 pages. <https://doi.org/10.1145/3274442>

## 1 INTRODUCTION

The study of computer-supported cooperative work in scientific research settings is an important subdomain of CSCW, particularly as more and more disciplines adopt computational approaches to their work [Jirotko et al. 2013]. Prior

---

Authors’ addresses: Andrea K. Thomer, School of Information, University of Michigan, 105 S. State St. Ann Arbor, MI, 48104, USA, athomer@umich.edu; Michael B. Twidale, School of Information Sciences, University of Illinois at Urbana-Champaign, 501 E. Daniel St. Champaign, IL, 61820, USA, twidale@illinois.edu; Matthew J. Yoder, Illinois Natural History Survey, University of Illinois at Urbana-Champaign, 1816 S. Oak St. Champaign, IL, 61820, USA, mjyoder@illinois.edu.

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM.

Manuscript submitted to ACM

Manuscript submitted to ACM

53 research in CSCW has explored how scientists collaborate at a distance [Li et al. 2012; Sonnenwald 2003], share data  
54 and software [Chin and Lansing 2004; Faniel and Jacobsen 2010; Pasquetto et al. 2016; Velden 2013; Vertesi and Dourish  
55 2011], and build and maintain infrastructures [Karasti et al. 2010; Karasti and Blomberg 2017; Korn et al. 2017; Ribes  
56 2014; Ribes and Finholt 2009]. Much of this research focuses on somewhat constrained instances of cooperative work:  
57 cases in which researchers are working together as part of the same project, organization, or well-defined collaboration.  
58 Fewer have explored the processes underlying community-scale CSCW in science, in which many researchers work  
59 toward a common goal, but at an “arm’s length” distance from one another — that is, “without direct communication  
60 and without necessarily knowing each other or knowing of each other, via a more or less shared information space...  
61 comprising data, personal beliefs, shared concepts, professional heuristics etc.” [Bannon and Schmidt 1989].  
62

63 While this sort of distributed-yet-cooperative work occurs in many disciplines, it is particularly fundamental to  
64 fields seeking to build knowledge organization infrastructures such as classification systems or ontologies mapping the  
65 composition of, and relationships between, different phenomena and entities. Disciplines such as biological taxonomy,  
66 certain branches of chemistry, and library and information science aim to collectively produce shared systems of  
67 classification [Price 1965]; other fields, such as biomedicine and ecology, necessarily draw on these classification systems  
68 as key infrastructure that facilitates other work [Hoehndorf et al. 2015]. Though some of these classification systems are  
69 managed by governing bodies, they are nevertheless contributed to by individual researchers working independently.  
70 Researchers must cooperate across distances of space (in that researchers are rarely co-located), awareness (in that  
71 they are often working independently and without direct communication), and even time (in that they must integrate  
72 new contributions with sometimes very old prior work). When adding to or revising a classification, researchers must  
73 reference and re-write existing classifications. The complexity of this work is only increasing as research communities  
74 move from relatively slow, paper-based publishing paradigms, to more rapid, granularized, and born-digital publications,  
75 in which research products include not just papers but also “nanopublications” and semantically-rich datasets (e.g.  
76 [Kuhn et al. 2016; Mons and Velterop 2009]).  
77

78 Here we describe a powerful example of “arm’s length” cooperative scientific work: the field of biological taxonomy.  
79 One of the oldest and most foundational sciences, taxonomy is concerned with the description, naming and organization  
80 of life on earth. Taxonomists fit the above definition of arm’s length cooperative work in that they are often geographically  
81 distributed from one another, as well as physically far from the habitats and creatures that they study. In some ways,  
82 they also work across large distances of time: not only do they sometimes work on projects that take years to complete,  
83 but they must integrate their data with prior work that is sometimes hundreds of years old.  
84

85 Through qualitative and design-based research methods, we show how taxonomists cooperatively-but-independently  
86 work to create a knowledge organization system documenting earth’s biodiversity. Our goal is to contribute to  
87 scholarship in CSCW, information science and biodiversity informatics while also designing better interfaces and tools  
88 for taxonomic work. We ask: What mechanisms and strategies do taxonomists use to cooperatively maintain and update  
89 systems of taxonomic nomenclature and description? What is the character of cooperative work in taxonomy, and what  
90 can that tell us about CSCW in general?  
91

92 We identify several characteristics of taxonomy that make its arm’s length cooperative work possible, and particularly  
93 identify *semantic refactoring* as a unique and important kind of articulation work [Gerson and Star 1986; Strauss  
94 1985] entailed in taxonomic maintenance. Given the longevity of taxonomic data, our findings have implications for  
95 infrastructure studies and the study of long-term, “slow” scientific endeavors [Stengers 2018], as well as to scholarship  
96 regarding the navigation, surfacing and maintenance, and reconciliation of classification systems. We conclude by  
97 outlining implications for the design of systems to support arm’s length cooperative classification work.  
98

## 2 “ARM’S LENGTH” COOPERATIVE WORK AND CSCW IN SCIENTIFIC SETTINGS

As briefly reviewed above, there is a substantial literature within CSCW on cooperative scientific research. Jirotko et al. note the challenges of studying such settings, as well as the challenges of understanding what may be needed for such collaborations to thrive:

The development of e-Science technologies and infrastructure requires methods of studying collaborations that are often diffuse, geographically distributed, asynchronous, and sometimes without clear delineations between users and developers [Jirotko et al. 2013].

“Arm’s length” cooperative work [Bannon and Schmidt 1989] is one such type of diffusion and distribution. It is distinct from other kinds of asynchronous collaboration in that it is extremely *loosely coupled*; workers may not even know each other, but nevertheless work toward a common goal. Work is often facilitated through the use of coordinative artifacts that multiple actors interact “through,” such as documents placed in locations that others can see and access, or specimens cataloged in a collection for common use [Schmidt and Wagner 2004]. The division of labor is typically very clear and broken down into discrete units, thereby enabling people to work relatively independently [Bannon and Schmidt 1989]. Via these strategies, workers are thereby able to keep down the well-known costs of CSCW — such as coordination, articulation work and maintaining awareness [Olson and Olson 2000; Schmidt 2002] — but do not, of course, reduce these costs to zero.

“Arm’s length” cooperative work is found in many settings; in their study of the coordinative processes in architectural design, Schmidt and Wagner declare that it is “*the salient characteristic of contemporary cooperative work*” [Schmidt and Wagner 2004] (emphasis added). However, we believe that it may be particularly salient to certain kinds of *scientific* cooperative work, especially long-term and infrastructuring endeavors. Where some studies have shown how time matters in fast-paced environments such as hospitals [Reddy et al. 2006; Sarcevic and Burd 2009] and paper mills [Auramäki et al. 1996], numerous studies of scientific collaboration have shown how slower paced, long-lived projects exhibit a different kind of time sensitivity. Karasti and co-authors call out “infrastructure time” as a temporal orientation that must exist in synergy with shorter-term “project time” scales in order to [Karasti et al. 2010]; others note that long-term projects entail different coordination and knowledge transfer practices than fast-paced environments, requiring ongoing stakeholder enrollment [Randall et al. 2015; Ribes and Finholt 2009], “anticipation work” to bring planned work into being [Steinhardt and Jackson 2015], and on-going data curation and knowledge management [Ackerman et al. 2013; Karasti et al. 2006]. In long-term projects, then, the coordinative artifacts — the records, specimens, artifacts, publications, protocols and lab books — that facilitate cooperative work perhaps take on additional stress, creating a need for more deliberate knowledge management and data curation processes.

### 2.1 Computer-supported cooperative classification work

Issues of infrastructuring are therefore closely tied to issues of knowledge management; this is especially the case when the infrastructure in question is *for* information or knowledge management. Artifacts like databases, collections, and data and metadata standards occupy a gray space between infrastructures and tools — they are both organizing structures and artifacts in and of themselves. This duality has been explored in CSCW and related fields to an extent (e.g. [Bietz and Lee 2012; Bowker 2000; Bowker and Star 2000; Hine 2006; Randall et al. 2011; Volda et al. 2011] among others), as have issues of knowledge management from a CSCW perspective (e.g. [Ackerman et al. 2013]) however we believe there is a growing need for research in this area, especially focusing on the cooperative maintenance and creation of complex knowledge organization structures.

157 By knowledge organization structures, we mean artifacts such as:

- 158 • *taxonomies*, in both the general and strictly biological sense: systems of classification and the principles that
- 159 underlie those classifications [noa 2018]).
- 160 • *controlled vocabularies*: an “organized arrangement” of terms created by a community to promote consistent use
- 161 across records or data sources, and thereby facilitate information retrieval [Harpring 2010];
- 162 • and *ontologies*: a set of concepts and categories in a subject area or domain that describes their properties as
- 163 well as the *relationships* between different concepts; when serialized in a machine readable format such as the
- 164 Ontology Web Language (OWL), these relationships can be leveraged to support advanced information retrieval
- 165 and inference [Guarino and Welty 2004]
- 166
- 167
- 168

169 These knowledge organization structures are not mutually exclusive; a taxonomy might make use of a controlled

170 vocabulary in describing its terms, and a taxonomy might be serialized as an ontology. For instance, while the Periodic

171 Table of the Elements might be most familiar in its eponymously tabular form, it can also be expressed as a simple list,

172 a database, or an ontology in which the relationships between elements and their parts are explicitly expressed [Cook

173 2002]. Similarly, while another famous classification system, the Library of Congress Subject Headings (LCSH) has

174 historically been published as a series of large, red, leather-bound tomes containing lists of terms, it can now also be

175 accessed through the Library of Congress’ linked data service, which encodes the terms in a slightly more ontology-like

176 fashion by including information about relationships between terms and groups of terms [Congress [n. d.]].

177

178

179 The focus of this study — biological taxonomy — is one kind of scientific taxonomy. Biological taxonomies are

180 first and foremost unique because, in addition to organizing organisms for search and retrieval, they are meant to

181 model the system of biological life as it has evolved over time. The methods of creating and maintaining the taxonomy

182 therefore change along with biologists’ methods of discerning evolutionary relationships. Thus, where one might

183 think that biological taxonomies are somewhat “settled” or unchanging, they are actually constantly being updated,

184 as taxonomists discover new species or revise prior species descriptions in keeping with new methodological norms.

185 Biological taxonomies are additionally distinct from other kinds of taxonomies because individual components (species

186 descriptions) of the taxonomy are published in a somewhat piecemeal fashion. Taxonomists do not wait until they

187 have described all of life on earth to publish, but instead, publish species descriptions as they are written. There are

188 strict rules from governing organizations such as the International Code of Zoological Nomenclature (ICZN) and the

189 The International Code of Nomenclature for algae, fungi, and plants (ICN) for what counts as a legal scientific name,

190 but otherwise the work is carried out by researchers working somewhat independently. The unfinished taxonomy

191 still functions as vital infrastructure for biological and ecological studies, even though it is perpetually being edited.

192 Finally, biological taxonomy is unique for its longevity: the system of biological nomenclature has been maintained for

193 centuries, and is likely the oldest knowledge organization system on earth.

194

195

196

197

198 Knowledge organization systems, including biological taxonomy, are by no means unstudied. Indeed, classification

199 research is foundational to library and information science. However, much of the existing research does not focus

200 on the day-to-day work of creating or maintaining a classification system, cooperatively or otherwise. For instance,

201 while Bowker and Star’s seminal book, “Sorting Things Out” [Bowker and Star 2000] does discuss the work of creating

202 classifications, this issue is somewhat tangential to the authors’ main arguments. Several other researchers have similarly

203 considered classification systems as boundary objects (e.g. [Albrechtsen and Jacob 1998; Jansen 2013; Rotman et al. 2012;

204 Star and Griesemer 1989]); but again, they do not necessarily focus on the work of creating the classification system

205 itself, but rather, how others use the classification system to do other work. As we noted above, the field of library and

206

207

208

information science has extensively considered the work of classification maintenance, examining work-centered (e.g. [Albrechtsen and Pejtersen 2003]), user-centered (e.g. [Losee 2006]) or warrant-based (e.g. [Bullard 2016]) approaches to developing classifications (among many others that are unfortunately beyond the scope of this paper to review). We wish to build on this work from a CSCW perspective, and take cooperative maintenance and creation of classification systems – particularly, over the long term – as our central concern.

We do note that there is a growing body of CSCW and HCI research around the collaborative and ongoing development of ontologies. Ontologies such as the Gene Ontology and the Foundational Model of Anatomy Ontology have become valuable infrastructure for biomedical informatics, and as such, have merited considerable investments in development and maintenance in recent years [Hoehndorf et al. 2015]. These ontologies are often maintained and used by globally distributed communities; and recent work has begun exploring the work needed to use and maintain these systems over time. For instance: [Vigo et al. 2014] identify key design insights for ontology development tools; [Vigo et al. 2015; Warren et al. 2014], outline activity patterns in the ontology authoring process. [Sarasua et al. 2015] point to the cooperative ontological work entailed in projects like WikiData; and [Breitenfeld et al. 2018] discuss the need to enable structured data creation by non-technical experts. However, we echo Randall et al.’s finding that that “modeling approaches to ontology building tell us little about the practical organization of the work and how this relates to the prospect of successful sharing” and that further ethnographic work is needed to understand these processes in more detail [Randall et al. 2011].

## 2.2 Prior studies of biological taxonomy

While scholars in CSCW and related fields have previously studied different aspects of biological nomenclature and classification, few have focused on the day-to-day work of workers in the discipline. For instance, Bowker considers the role of biological taxonomies as scientific artifacts [Bowker 2000], as well as the conflicts that arise when communities use different terms for the “same” species [Bowker 1999] but does not focus on the work of reconciling these differences. Rotman et al. consider biological taxonomies as boundary objects, and while they discuss the cooperative work taxonomies facilitate, they do not discuss the cooperative work put into their creation and maintenance [Rotman et al. 2012]. Hine’s study of taxonomy as a cyberscience focuses on their adoption of information technologies for communication, but not necessarily their day-to-day work practices [Hine 2008]. Recent work by Montoya and Erickson considers the process of taxonomic reconciliation by discussing the consequences of taxonomies that *don’t* get updated in the online resource, the Catalog of Life [Montoya and Erickson 2017], but while this is a valuable analysis, the goal of the study simply is not to focus on taxonomists’ work practices. This prior work lays an important theoretical foundation for the study of taxonomic work practices, but further research is needed to describe and support day-to-day taxonomic work.

The bulk of the work on taxonomic work practices has actually been written by taxonomists themselves. There is a robust literature considering everything from the population dynamics of the field [Joppa et al. 2011], to the implications of a decreasing rate of new species descriptions written per taxonomist [Costello et al. 2013], to analyses of the revision rate of taxonomies over time [Vaidya 2017], to the logical consequences and limits of applying biological names [Franz et al. 2016]. Taxonomists have broadly acknowledged a need for modernized software, and modernized modes of data publishing; many have particularly emphasized the need to use ontologies and other semantic web technologies to make their data more machine readable and interoperable [Balhoff et al. 2013; Cui 2010, 2012; Deans et al. 2012; Yoder et al. 2010, 2018]. However, further research is to understand how best to support the complex work of taxonomy, not to

261 mention the complex work on ontology development and application. Here we intend to contribute to the construction  
262 of that bridge, by discussing taxonomic work in a CSCW context.  
263

### 264 **3 APPROACH** 265

266 Our research team includes information scientists, taxonomists, and software developers interested in understanding  
267 and supporting the cooperative work of taxonomy development, use and management. Over the course of our four-year  
268 project, we iterated between qualitative and design-based methods of research. This approach allowed us to conduct  
269 research while also building community and improving existing tools. We summarize our data collection approaches,  
270 as well as our methods of analysis, below.  
271  
272

#### 273 **3.1 Qualitative research methods: site visits and semi-structured interviews** 274

275 Our study is grounded in qualitative research methods. We conducted ethnographic site visits and semi-structured  
276 interviews with taxonomists in order to develop a detailed model of taxonomists' work practices, software preferences,  
277 and unmet needs. From 2014-2015, we visited five different sites and interviewed a total of thirty-four taxonomists.  
278 Participants were initially recruited from collaborators on our funding grant, and then snowball sampled from that  
279 group to include other "nearby" researchers, and to those using the taxonomic software<sup>1</sup> produced internally by us and  
280 our parent organizations; consequently, our participant demographics largely reflect the users of this software, and  
281 are consequently skewed toward entomologists (particularly hymenopterists, who study wasps). Twenty-eight of our  
282 participants worked in entomology; two in botany; two in mycology; one in invertebrate marine biology; and one in  
283 invertebrate paleontology. However, given that entomology is one of the primary disciplines engaged in taxonomic  
284 description, we believe that they make an excellent user community to ground this research in. Only seven of our  
285 thirty-four taxonomists identified as female; this may also reflect the demographics of the taxonomic community  
286 [Jackson 2015].  
287

288 Site visits lasted three to five days each. During visits, we observed participants while they worked; sat in on  
289 meetings; and photographed their workplaces. We drew on methods of rapid ethnography [Millen 2000], which involves  
290 a constrained research focus, using key informants, and capturing rich field data using multiple perspectives. In semi-  
291 structured interviews, we asked participants to demonstrate and describe their workflows in greater detail. We also  
292 asked them questions about their data management practices; decision making strategies; preferences in hardware  
293 and software; frustrations and successes in taxonomic data creation and management; and other details about their  
294 day-to-day work (see Appendix A for our interview script and site visit guidelines).  
295  
296  
297  
298  
299

#### 300 **3.2 Design-based research methods: rapid prototyping and design hackathons** 301

302 In addition to our qualitative data collection, we drew on design-based methods of research, including participatory  
303 design and rapid prototyping. Our research team met weekly and reviewed notes and memos from site visits as they  
304 were completed. We developed a preliminary conceptual model of taxonomists' work practices and common information  
305 interactions, as well as a rough list of issues they encountered in their work. We then brainstormed interface ideas  
306 that might better support their needs; we additionally used the process of designing interfaces to refine our conceptual  
307 model of work practices.  
308  
309

310 \_\_\_\_\_  
311 <sup>1</sup><http://taxonworks.org/>



313  
314 After completing all site visits, we held a three-day Interface Design Hackathon. The goal of the Hackathon was to  
315 further refine our initial conceptions of taxonomists’ work practices, while also directly involving taxonomists in our  
316 design work (described previously in [Thomer et al. 2016; Yoder et al. 2018]). The Hackathon was also intended to build  
317 community and encourage on-going collaboration between our team and taxonomists. Given that much of taxonomic  
318 software is developed and maintained by taxonomists themselves, we felt it was important to involve them as directly  
319 as possible in our work. Hackathons have become a common method of software development in the biosciences (see  
320 [Katayama et al. 2010; Lapp et al. 2007]), and have been shown to provide participants a means to identify problems,  
321 define community goals and priorities, and share tutorials and technical knowledge. As such, this was an appropriate  
322 format for collaboration with this community.  
323

324  
325 Twelve participants attended in total, including four project staff, six taxonomists who had previously participated  
326 in interviews or site visits, and three taxonomists from sites we had not visited. In planning the Hackathon, we drew on  
327 methods of participatory design [Blomberg and Karasti 2012]. We asked participants to complete three-part workbooks,  
328 designed to capture specific details about their work practices as well as their interface designs. The workbooks consisted  
329 of: 1) a use case template; 2) a flowchart template; and 3) a space for wireframe interfaces (see Appendix B for the full  
330 workbook template, and Figure 1 for a small excerpt from a completed workbook). The workbooks were hosted on  
331 Google Docs, and participants were pointed to flowcharting and wireframe design tools, such as LucidCharts, Balsamiq  
332 and VUE. We also provided participants with materials for paper prototyping.  
333

334  
335 Teams self-organized to fill out workbooks. Sixteen workbooks were completed by the end of the Hackathon, and  
336 four more were initiated but not completed. On the last day, participants voted on which of the use cases they would  
337 like to see developed first. Paper prototypes were digitally photographed, and inserted into the digital workbooks.  
338

339 After the Hackathon, our project team resumed our weekly-to-biweekly prototyping meetings to further review  
340 notes, and to iterate on designs generated through the Hackathon. We translated and refined these paper prototypes  
341 into formal wireframes, and worked with software developers to build prototypes of these interfaces in our testbed  
342 software environment. Approximately three designs were integrated into working prototypes, implemented in this  
343 project’s associated software environment and further refined through user testing [Yoder et al. 2018].  
344  
345  
346  
347

### 348 **3.3 Analytic approach**

349  
350 In keeping with our rapid ethnographic approach, much of our preliminary analysis was conducted through our research  
351 team’s meetings over the course of this project; as described above, this allowed us to develop a heuristic, yet grounded,  
352 understanding of our participants’ work practices. By iterating between analysis, data collection, and design work, we  
353 were able to pivot quickly to test or refine new ideas and insights as they emerged.  
354

355 We followed this heuristic evaluation with a more structured analysis of interview transcripts and ethnographic  
356 memos. After all interviews and site visits were completed, we used Dedoose to code notes from interview transcripts  
357 and notes from site visits. While we largely took a grounded approach in this work (coding transcripts for emergent  
358 themes, which were then refined through discussion and comparison of codes [Charmaz 2014; Glaser and Strauss 2017]),  
359 we also paid particular attention to themes that emerged during our preliminary analysis, such as issues of workflow,  
360 data and concept management. Transcripts and notes were coded by all three authors of this paper; conflicting codes  
361 were discussed in project meetings and brought into accord.  
362  
363  
364

## 4 RESULTS: CHARACTERIZING THE WORK OF TAXONOMIC WORK

Derek De Solla Price placed biological taxonomy into the category of sciences producing a body of knowledge, and a system of classification that is “optimistically considered more or less eternal” [Price 1965]. We found that though this view of taxonomy is not necessarily wrong, it is also not quite complete. The researchers we spoke with were quick to note that while a broad goal of taxonomy is indeed to develop a collectively agreed upon Book of Life, the emphasis is increasingly on a *plurality* of books - rather than one monograph to rule them all. There are necessarily many ways to identify and constrain what counts as a species, often depending on the type of organism – describing a microbe requires a different set of methods, tools and terms than describing a tree – as well as the resources a taxonomist has at their disposal – some labs may not have the equipment or staffing to support genetic sequencing. Therefore, there are necessarily many ways to classify species and their relationships between each other, which must be reconciled.

Through our site visits, interviews and design-based engagement with taxonomists, we refined the following account of the work of taxonomy. We first describe the iterative process of refining and revising taxonomic descriptions. Then we describe how participants come to learn these data practices through various enculturation practices. Throughout, we present a selection of interfaces designed by us and our participants to underscore the kinds of information interactions entailed in taxonomic work.

### 4.1 Writing taxonomic descriptions: iteration and revision

The fundamental task in taxonomy is to create species descriptions: semi-structured (referred to as “telegraphic”) narratives describing what makes a group of organisms unique. For instance, in the following example from a description of a kind of near-microscopic wasp, the author describes details of the animal’s teeth, wings, wing hairs (*setae*), as well as differences in the shape of the tail end (*gaster*) in male and female specimens:

Subfamily TETRACNEMINAE. Mandibles bidentate; forewing without a filum spinosum; setae on basal cell of similar size to those beneath apex of venation.... FEMALE - gaster with last tergite more or less shield-shaped or triangular, its anterior margin almost straight, hardly curved [Noyes 1988]

The “semi-structured” aspect of species descriptions refers to both their form and content. In terms of their form, descriptions are often organized consistently, listing anatomically equivalent elements in the same order from one description to the next. In terms of content, taxonomists typically seek to use controlled vocabularies that are recognizable to their peer group (the tensions around selecting controlled vocabularies are described further below). Species descriptions can be published individually, but more often multiple descriptions are published in one article or monograph (see the journals *Zootaxa*<sup>2</sup> and *Zookeys*<sup>3</sup> for examples of these publications). Descriptions for very large groups of organisms may even be published in a series of books. For instance, one participant is currently in the midst of publishing an anticipated *ten volume* series on a hyperdiverse group of chalcid wasps, which he anticipates will take *ten years* to write.

Descriptions are written through a highly iterative process of collecting specimens (both from the wild and from existing museum collections) and identifying diagnostic *characters* (aspects of an organism’s anatomy) that distinguish one group from another. Just as the presence of fur helps to distinguish mammals from reptiles and birds (Table 1), the presence of a mostly straight anterior edge of the last segment of an insect’s “gaster” (tail end) helps to distinguish female from male *Tetracneminae*. Descriptions are sometimes created along with, or via the proxy of a character matrix.

<sup>2</sup><http://www.mapress.com/j/zt/>

<sup>3</sup><https://zookeys.pensoft.net>



Table 1. A very simple character matrix. Each **X** notes the presence of a “character” in an organism. These matrices are used to begin working toward a hierarchical classification of organisms into groups. In this case, the taxonomist would need to consider whether a Platypus would be better grouped with Lions and Zebras rather than its fellow egg-layers, the Lizards.

Taxon	has 5 fingers	has fur	lays eggs
<i>Lion</i>	X	X	
<i>Lizard</i>	X		X
<i>Platypus</i>	X	X	X
<i>Zebra</i>		X	

### Simple Matrix Editing with considerations for character exploration

**Background & Goals**

- We want a matrix editor that has minimal barriers to start use, but also can draw from existing records when available. It should be able to facilitate character exploration and communication when characters haven't been formalized/formally coded yet.
- Considerations for communicating should include teaching/training as well as collaboration with other people. Facilitates a simultaneous user interface.
- Concept is a "google doc your matrix", the matrix is not just for one person to get.
- The core tool is part of a larger workflow (see approach)

**Workflow/Approach (Relevant Use Cases)**

- Land on the matrix
  - Rows are added by
    - ID of prep
    - name of OTU
    - tagged set of preps/otu
  - Columns are added by
    - ID of character
    - name of character
    - tagged set of characters
  - For Rows, if no match on search option to add
    - OTU
    - prep/specimen
  - For Columns, if no match on search option to add
    - new exploration/brainstorm (EB) character

Fig. 1. An excerpt from one of our Hackathon use case notebooks describing a proposed Matrix Editing interface. In this interface, each data point in a character matrix (such as the X’s in Table 1) would be hyperlinked to supporting data files, thereby improving the reproducibility of a taxonomic description.

These matrices have species or taxa as rows, and characters (sets of related states) as columns. Individual cells are filled with data (i.e. the taxonomist’s observations). In our design work, we and our participants imagined several interfaces for character identification and definition, as well as character matrix manipulation and design. We found a particular need for tools that captured provenance, such as links between various data sources, and thereby recorded why exactly the author considered one wasp’s tergite more or less triangular than another’s. One design from the Hackathon included a tool for “Simple Matrix Editing with considerations for character exploration,” and emphasized the need for a Google-doc-like level of collaborative capacity, combined with a database-like level of data management and storage (Figure 1).

4.1.1 *Iterative workflows.* The process of identifying and classifying characters varies from taxonomist to taxonomist; we observed considerable differences in our participants’ workflows, and in interviews, they described their field’s practices as being quite diverse. Some take notes on potential species groups in the field, as they gather specimens;

469 others wait until they are back in the lab. Some first group specimens by eye, and then write down key characters;  
470 others take a more structured approach and survey specimens one-by-one and fill in a character matrix as they go.  
471 However, despite this variation, every taxonomist we spoke with agreed that the character definition process was  
472 necessarily iterative. In their words:  
473

474 I do a minimum of two or three passes for specimens. First one is to get an idea. The second one is to look  
475 at the limits, then I test my original hypothesis with more data. (J-1-TOR)

477 I'll finish something, and I might get more material about what I think I've finished. I take [the specimens]  
478 back and go through the same process again of what I've written before and see if they come out okay.  
479 (N-3-NHML)

481 You reassess the characters so you go through them. We all have a notion - well, I have a notion - of which  
482 things should fit together when you look at your specimens or your species... And then you run your  
483 analysis and if everything comes out crazy then you have to reassess; you have to go back and look at it  
484 again because you have to make sure that it's crazy and it's not just bad results. Yeah, you go through it  
485 again until you're sick of it basically. (D-2-NHML)

488 Sometimes in reading literature you find another author described a character you didn't see, so you go  
489 back to see if it's there or if you missed it. (R-1-IL)

490 Character definitions and species concepts are thus something that emerge from repeated sorting and review. This  
491 process is not unlike that of developing a grounded theory, with the caveat that our taxonomists may be somewhat  
492 more abductive in their approach than grounded theory prescribes: they bring their training and prior experience with  
493 a group of organisms to bear on their classifications. When they talk about assessing whether data are "okay" and not  
494 just "bad results," they are drawing on their prior experience with this group and dataset to make a judgment call about  
495 the quality of the analysis.  
496

498 We noted that taxonomic work in general, but species description workflows in particular, have a very visible  
499 and inherent materiality to them. Physical specimens are the primary data source, and no amount of photography  
500 or digitization will replace them. Thus, taxonomy is certainly computer supported, but not by any means computer  
501 dominated; and taxonomists' work must bridge the digital-physical divide by iterating between a wide range of data  
502 sources and tools: physical specimens, and physical photographs, microscope slides, paper grids, old books, digital  
503 images, DNA sequences and databases (Figure 2). Our design work focused on brainstorming ways of making this  
504 bridge a bit more seamless; for instance, creating labels that were both human- and machine-readable, or lightbox-like  
505 interfaces that would allow taxonomists to group and categorize their digital data as easily as their physical (Figure 3).  
506  
507  
508

509 *4.1.2 Iterative Revisions.* Taxonomists are also engaged in a longer-term kind of iteration: the review and revision  
510 of prior taxonomies. Participants emphasized that each description that they write is a hypothesis, which will be  
511 tested and re-tested by future researchers. Revisions are necessary when older descriptions are not clear, thorough, or  
512 otherwise complete enough to be usable in modern research. For example, several participants noted that they were  
513 working on "molecular revisions" of groups, in which DNA sequences are taken from specimens and used in addition  
514 to morphological characters to revise a taxonomy.  
515

516 In other cases, revisions were motivated by the discovery of new organisms, and prior classifications had to be altered  
517 to include them. The process of defining characters involves not just looking at specimens, but also at prior descriptions  
518 of specimens, and then making a decision about what characterizations to prioritize. One participant described this as  
519

521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572



Fig. 2. A taxonomist’s workspace, including a computer, books, specimens, a microscope and many other materials. Photograph by Chris Nixon.

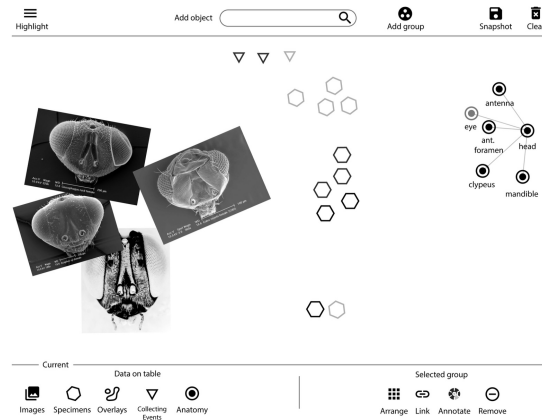


Fig. 3. A virtual light table. The table space include images, and symbolic representations of specimens, collecting events, and an anatomy ontology arranged in a freeform open space. Basic functionality includes adding, highlighting, selecting, manipulating, and annotating the current contents of the table (bottom). Previously published in [Yoder et al. 2018]

forcing the revising author to decide what prior work and prior data to carry forward: if “one description gives shape, color, and shell thickness, and the other gives you description of shell sculpture and hinge teeth, like, ‘it is heavily sculptured and the hinge has three teeth’ ” (R-1-IL) the revising author would need to decide which (if any) of those characters to use in the revised descriptions.

#### 4.2 Enculturation in data practices: selectively controlling vocabularies, and passing down tools

When asked about their workflows, our participants inevitably described their training. Participants described a strong tradition of apprenticeship in taxonomy, through which junior taxonomists are enculturated in their mentors’ data practices. During site visits and in our Design Hackathon, several participants said that they used the workflows, techniques and tools that they did at least in part because their mentors trained them that way, and fondly described

573 their mentoring relationships and connections of practices in different labs. One participant said that this may be part  
574 of the reason behind the proliferation of idiosyncratic methods and software preferences in taxonomy, noting that,  
575 “Many people are using tools that were handed down by their predecessor or that they designed themselves” (R-1-IL).  
576

577 The extremely loose coupling of the collective endeavor of taxonomic classification permits these varying schools of  
578 workflows and personal idiosyncrasies. Provided the product (a species description) follows established forms, consider-  
579 able variation in the process (a taxonomist’s workflow) is tolerable without causing much disruption. Consequently, the  
580 collective result is robust under substantial diversity of practice, again contrasting with smaller, more tightly coupled  
581 scientific collaborations. For taxonomy, this is both a challenge to widespread systematic adoption of new practices and  
582 technologies, but also an opportunity for small scale innovation.  
583

584 As we described above, in writing a description taxonomists must describe anatomical characters as clearly and  
585 unambiguously as possible, so that others can use them to identify specimens or determine whether they have discovered  
586 a new species. However, taxonomic data’s extremely long lifespan is in itself an obstacle: the descriptions can last longer  
587 than our common understanding of the terms used to write them. Though individual taxonomists try to use preexisting  
588 vocabularies when possible, often they are too arcane to be clarifying. For instance: the terms “tomentose”, “floccose,”  
589 “arachnoid,” “hoary,” and “lanate” all refer to kinds of fuzziness that might be seen on a leaf [Armstrong 1999] – yet,  
590 their use may confuse more than inform a modern reader. Thus, while individual taxonomists strive for consistency,  
591 they still must make hundreds of ad hoc decisions about which standards to adhere to, and which to ignore. As one  
592 botanist explained:  
593  
594  
595

596 We have some standard references that we have, standard botanical Latin. That’s a synthesis of what  
597 Linnaeus used and what’s out there from way back. It gives you the roots, but it’s only useful if everyone’s  
598 using that same reference. The problem is they’re not. Leaf shape can be terrible. Someone’s “ovate” is the  
599 other person’s “elliptic” or “lanceolate.” Those characters, some people don’t use fancy labels because they  
600 don’t believe it’s distinct from whatever [word] they just imagine. [M-1-NHML]  
601  
602

603 Participants also explained that they chose terms based on the preferences of immediate colleagues. As one summarized:

604  
605 If other people have used other terms that I’m not used to... if they’re good, I will take them. But if all my  
606 colleagues use a different term [than other people], those are the people I’m communicating with, so I  
607 will probably use the term that [we’re all] familiar with. (D-2-NHML)  
608

609 Thus, vocabulary selection is partially driven by the need to create descriptions that clearly express concepts and match  
610 historic terms, but is also driven by each taxonomist’s in-group norms and specific training. Taxonomists select terms  
611 based on both historical precedent and their own best judgment about the precision and interpretability of a given word  
612 for its intended audience; in doing so, they effectively modernize the vocabulary of taxonomy on a case-by-case basis.  
613  
614

615 *4.2.1 Using images to bridge differences in vocabulary use.* Given the unquantifiable nature of many anatomical  
616 characters (e.g. “slightly curved,” “fairly sculpted,” and so on), images are critically important to describing diagnostic  
617 anatomical characters, and are often used to help bridge the gap between different terminological cultures. As one  
618 participant (who works on moths) explained, images are important in clarifying how terminology is being used by  
619 researchers with different backgrounds:  
620

621 When we get together and we talk about wing patterns, we’ve been brought up on different literature ...

622 There are differences in how we assign the terminology that we use so we will have to either use images  
623

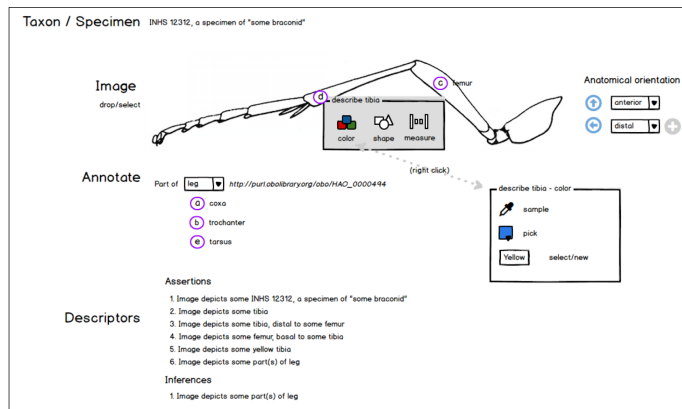


Fig. 4. A wire frame of an interface for image annotation. A user would use the cursor to select some point of the image, and right click to bring up a menu of tagging options. In this example, an image of an insect leg is being annotated with descriptions of its color; the specific terms being used are drawn from an underlying ontology.

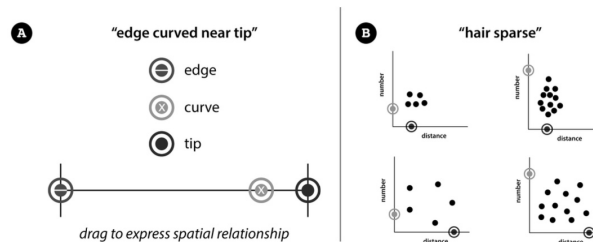


Fig. 5. A wire frame of an interface for to add visual annotations to qualitative terms. The taxonomist would use the sliders above to create simple images to clarify their use of fuzzy terms. Previously published in [Yoder et al. 2018]

with drawings on top and say this is what I mean because the characters have to be consistent when we score them because if not, they make no sense in the end. (D-1-NHML)

Despite their utility, the ability to rely on images rather than (or in addition to) text as the primary mode of scholarly communication is fairly new; there have historically been limits on the number of images that could be included in a paper-based publication either due to page constraints or the high cost of color printing, which are now being lifted thanks to the advent of electronic publishing. Images used in taxonomy range from photographs, micro-photographs, and hand drawn and computer-aided illustrations of specific anatomical features over the course of their work. The specific kind of illustration used often depends on the qualities of the organisms being discovered: very small insects who can only be determined through characters found on internal organs or genitalia will need to be dissected, slide mounted, and photographed through a scanning electron microscope; whereas larger organisms can simply be mounted and photographed with a standard digital SLR.

In our design work, we and our participants imagined interfaces that would support the semantic annotation of images (Figure 4). Users would import an image, and drag-and-drop qualifiers to different aspects of the image. These terms would be linked to anatomical ontologies; users would thereby be able to create semantically rich data, linked to images, without necessarily having to interact directly with an ontology. We additionally designed interfaces that could

677 add some kind of visual anchor to qualitative descriptors such as “sparse” or “curved,” thereby allowing taxonomists  
678 to continue using natural language while nevertheless ensuring that their meaning is communicated (Figure 5). This  
679 would make it possible to continue taking advantage of the qualitative richness of text, while reducing ambiguity (and  
680 improving machine readability) through a visual aid or quantitative measure.  
681

## 682 683 5 DISCUSSION 684

### 685 5.1 Taxonomy as an exemplar of arm’s length cooperative classification work 686

687 This project contributes a rich description of the practice of biological taxonomy, and several insights into classification  
688 work and maintenance with implications for CSCW and related fields.

689 First, we believe that biological taxonomy should be viewed as an exemplar of a successful large-scale infrastructure  
690 project. The process of naming and classifying organisms is one of humankind’s longest term endeavors; and the formal  
691 system of taxonomic nomenclature has been in continual development for nearly three hundred years. Taxonomy’s  
692 global and organizational scale encompasses the world community of taxonomists; even very large projects such as  
693 CERN or the Long Term Ecological Research Network cannot match that scale. As such, despite the challenges other  
694 authors have noted in using taxonomic classification systems [Rotman et al. 2012] or in keeping individual instances of  
695 nomenclature up-to-date [Montoya and Erickson 2017], the discipline of biological taxonomy nevertheless represents a  
696 major success story in infrastructure maintenance, and a potential counterpoint to studies of failure and breakdown.  
697 Taxonomists might be viewed as a class of “maintainers” [Russell and Vinsel 2016]: through their iterative workflows  
698 and on-going revision of the classification system, they ensure that the conceptual infrastructure that is taxonomic  
699 nomenclature stays up and running.  
700

701 Second, we find that the loosely-coupled cooperation of taxonomic work is facilitated by a number of factors: the  
702 clear divisions of labor within the field (in that researchers or research labs tend to focus on a relatively narrow range  
703 of taxa); strong apprenticeship and mentorship traditions; and the reliance on a common set of norms and standards for  
704 data *formats and quality*, but not necessarily workflows (for instance, describing anatomical characters in a regular  
705 order; adhering to guidelines for publication from the nomenclatural bodies). The standards that do exist are somewhat  
706 “loosely-coupled” as well; vocabularies are controlled, but not at the expense of expressiveness. This combination of  
707 community norms with a discipline-wide understanding that one must use one’s best judgment makes it possible for  
708 taxonomists to work alone or in small teams while contributing to a collective and growing whole. This community’s  
709 coordination and enculturation practices could inform other fields of cooperative knowledge structure or classification  
710 work, including crowdsourced work; citizen science; and scientific ontology manipulation and creation. These loosely  
711 coupled coordination practices may additionally be informative to communities engaged in more tightly managed  
712 classification and vocabulary management (for instance, metadata standards bodies, or other scientific communities  
713 who create and rely on controlled vocabularies).  
714

715 Third, the inherent and quite literal materiality of taxonomic work makes the community an excellent testbed for tools  
716 that broach the physical/digital divide. Though taxonomists image, measure, extract DNA from, and otherwise transform  
717 physical materials into digital data, the physical specimens must nevertheless be preserved for future examination  
718 and reuse. Taxonomists also iterate between physical data processing (grouping, arranging, observing, dissecting,  
719 labeling specimens) and digital data processing (defining characters, creating matrices, writing descriptions, annotating  
720 images). They largely link the physical/digital worlds through labels and identifiers, but there is opportunity to develop  
721 tools to make this jump between data manipulation modes more seamless. In our on-going work, we are exploring  
722



729  
730 the application of augmented reality interfaces to support taxonomic work, and hope to publish the results of these  
731 endeavors soon; we believe there is potential for innovative work in this area.

732  
733 Fourth, the kind of data that taxonomists produce is in many ways an ideal test case for the born-digital publication  
734 of linked data or “nanopublications” [Kuhn et al. 2016; Mons and Velterop 2009]. As demonstrated above, taxonomic  
735 descriptions entail the manipulation and creation of a wide array of controlled vocabularies, nomenclatural bodies, and  
736 provenance chains. In many ways, this data is already “Linked” — at least, conceptually — which has spurred many  
737 taxonomists to argue that they need to move to a linked data, semantically-annotated publishing paradigm [Balhoff  
738 et al. 2013; Cui 2010; Deans et al. 2012]. The taxonomic community may be an ideal use case for the development of  
739 tools for the management and creation of linked scientific data. Given that current methods of producing linked data,  
740 whether taxonomic or otherwise, tend to require fairly advanced knowledge of the encoding standards themselves,  
741 this approach has thus far been limited to people with extensive training in information and computer science. The  
742 interfaces we have developed have focused on capturing the semantics of taxonomists descriptions without forcing  
743 taxonomists to become experts in ontological application. We hope that the insights we have gained from this work can  
744 aid in the development of other tools to support the use of semantic technologies by non-semantic-technologists.

745  
746  
747  
748  
749 *5.1.1 Semantic refactoring: data migration through articulation work.* We wish to particularly highlight one aspect of  
750 taxonomic work that is key to both its longevity, and to the facilitation of taxonomists’ loosely coupled cooperative work.  
751 We see the process of revising classifications as a form of data migration. By revising older classifications, taxonomists  
752 migrate data from historic to modern formats. The practice of revising — rather than replacing — the classification  
753 makes it possible to build on prior contributions while also enabling the dataset of species descriptions to renew, to be  
754 sustainable, and to remain of current value.

755  
756 This revision process entails substantial articulation work. Here we mean “articulation” in two somewhat more  
757 literal senses than it is sometimes used: articulation as in enumerating — “the act of giving utterance or expression” to  
758 morphological characters and species concepts — as well as articulation as in “jointing or interrelating” these concepts  
759 into a classification or nomenclatural system [art 2018]. Both forms of articulation work require different kinds of  
760 information interactions, which we found to be sorely under-supported by current software. Articulation-as-in-defining-  
761 terms entails reviewing existing terms and their uses, consulting with community members, and selecting and defining  
762 new terms (or version of terms). Articulation-as-in-interrelating entails manipulation of species descriptions and  
763 conceptions; navigating historical classification and nomenclatural structures; and drawing relationships between these  
764 historical and new concepts in a human and ideally, machine, readable way.

765  
766 We might think of these unique kinds of articulation work as a process of *semantic refactoring*: the piecemeal practice  
767 of making semantic representations and structures clearer, more efficient, and more usable to a community. We borrow  
768 this term from computer science, in which it has previously been used to describe the process of replacing code with a  
769 clearer, but equivalent version (see for example [David et al. 2017]). *Here we use “semantic refactoring” to refer to the*  
770 *process of revising a controlled vocabulary or classification structure in response to changing norms or user needs.* The  
771 semantic refactoring entailed in revising a taxonomy requires updating the content of a species concept as well as the  
772 relationships between species concepts. Semantic refactoring makes it possible to carry forward the prior “codebase” of  
773 a classification system while correcting, amending, and building upon it. Though some of the processes described here  
774 are unique to taxonomy, we believe that semantic refactoring is likely central to other long-lived classification systems  
775 or ontologies. Updating subject headings in the LCSH, for instance, similarly involves reviewing and revising terms,  
776 their definitions, and their placement within a classificatory structure in response to changing use and users.

Table 2. Implications for design, summarized

Taxonomic work practice or characteristic	Implication for design
<i>Diversity in workflows</i>	Standardization should focus on data and publication formats but not workflows; quality rather than strict content standards.
<i>Loosely coupled cooperative work: independent researchers working toward a common goal</i>	Systems and workflows should be divided into discrete, well defined tasks.
<i>Semantic refactoring; need to reconcile conflicting classifications, or new and old</i>	Methods of visualizing and comparing taxonomies; version control and provenance tracking; collaborative ontology annotation tools; tools to review terms and definitions, tools to visualize history of concepts and their use.
<i>Bridging the digital/physical: work takes place both in and out of a computer</i>	Simple machine- and human-readable labels; augmented reality tools for data manipulation and management.

## 5.2 Implications for design

This project has generated several broad requirements – and indeed anti-requirements – of taxonomic work: functionalities that taxonomists do and do *not* want or would struggle to integrate into their work, and implications for design of those functionalities (Table 2). There is a clear need to account for the idiosyncrasies of taxonomists’ individual workflows, as influenced by the needs of the specimens being analyzed and the work styles inherited through their training. As found in numerous other settings, a one-size-fits-all solution is unlikely to be widely adopted. What is needed is a design for incremental adoption and support for substantial customization.

There was also great interest amongst taxonomists in designs that facilitate digitization and digital data curation, but with the caveat that these occur within an irreducibly material kind of information work. Specimens are central, and taxonomists are not in the business of digitizing the physical away. Instead, they aim to work toward a more born-digital work *process*, in which specimens can be represented digitally and annotated with images, text and literature reviews. They would also like to support pairing structured data representations (e.g. linked data) with free text descriptions, so that the resultant description can be computationally analyzed and reasoned over. Computers are not seen as a way of replacing existing workflows, but rather, augmenting them.

*5.2.1 Implications for other cooperative classification workers.* Findings from this study have implications for other non-biological and even non-scientific taxonomies and classification systems as well. First, as noted above, the loosely coupled nature of taxonomy is key to its longevity, and other domains of classification or knowledge organization work may wish to explore this model. While there are governing bodies that publish rules for what counts as a valid classification and name (e.g. the ICZN), classifications are not solely written by that body. While taxonomists are committed to using common terms and controlled vocabularies, they nevertheless understand that these need to be altered over time, and that boundary objects such as images and annotations are needed to truly communicate what is meant by a term. These practices may be adoptable by other communities.

833  
834 Further work on, and better tools for, semantic refactoring in a CSCW context is needed. There has been some  
835 research in this area, but more is needed. For instance, [Puente and Díaz 2012] look at using mindmaps in wiki refactors;  
836 Greenberg et al., discuss the need for a “community-driven metadata vocabulary application” that would support  
837 collaborative definition negotiation and term selection [Greenberg et al. 2013]; and Kunze et al. describe a system for  
838 supporting the evaluation of “digital stickiness” of a term or identifier — basically, a way of assessing the potential  
839 longevity of a resources [Kunze et al. 2017]. These approaches are promising but further workflows and workbenches  
840 are needed that tie these processes and resources together, and thereby better support the articulation work of semantic  
841 refactoring.  
842  
843

## 844 845 846 **6 CONCLUSION AND FUTURE WORK** 847

848 In this paper, we discussed biological taxonomy as exemplary of extremely long-term, cooperative scientific classification  
849 work. We contribute a rich description of taxonomists’ work practices, and the implications those practices have for  
850 design of cooperatively developed classification systems, both for taxonomists and for other knowledge organizers. We  
851 find that taxonomists rely on a unique kind of articulation work — semantic refactoring — to maintain and migrate  
852 their controlled vocabularies and classifications over time. We believe that semantic refactoring is also key to the  
853 maintenance of other classification systems, and that supporting this work will be critical in also supporting classificatory  
854 infrastructure. In our future work, we hope to further refine our understanding of taxonomic work practices through  
855 further design and testing of our interfaces; we additionally hope to compare taxonomic work practices to those of  
856 other cooperative classification communities (for instance, community-developed folksonomies, or the LCSH). We  
857 additionally hope to explore the role that taxonomists’ values, politics and priorities play in the decisions they make in  
858 their classification work; findings to this end were emergent in our data but out of scope for this paper.  
859

860 We conclude by noting that although infrastructure may be much easier to see when it breaks down [Star and  
861 Ruhleder 1996], taxonomy provides an example of its visibility through successful resilience and longevity. Schmidt and  
862 Bannon cite Neurath’s metaphor of scientific collaboration as “akin to sailors rebuilding their boat while at sea, without  
863 being able to start from scratch and without having recourse to an overarching plan” [Schmidt and Bannon 2013]. This  
864 is a most appropriate metaphor for taxonomic work. Having identified the proverbial Ship of Theseus, in future work,  
865 we hope to further understand how the sailors managed to rebuild their boat while at sea, and over so many decades.  
866  
867  
868  
869

## 870 871 **A SITE VISIT GUIDELINES AND INTERVIEW SCRIPTS** 872

### 873 **A.1 Observation guidelines for site visits** 874

- 875 • Do they setup a workflow of many set steps, or do the steps seem to be randomly accessed?
- 876 • What does the workspace look like (take pictures if allowed)? What items/tools are immediately near-at-hand?
- 877 • What tools (books, spreadsheets, computers etc.) are used to record observations?
- 878 • What tools are used to manipulate specimens being observed?
- 879 • How many steps are between the “aha” moment and the final representation of the anatomy/morphology being  
880 observed?
- 881 • Some work things out smoothly, some do not; particularly note the latter.
- 882 • What processes appear to be “linear” vs. “iterative”? Particularly note the latter  
883  
884

## 885 **A.2 Interview questions and prompts**

886 Background:

- 887
- 888 • Please tell me about what you study. What kinds of organisms? What characteristics?
- 889 • How many years of experience do you have?
- 890 • How did you learn to describe morphology/anatomy?
- 891 • How do you teach others to learn morphology/anatomy?
- 892 • How many species have you described?
- 893 • What year did you describe your first species?
- 894 • What year did you identify your first species?
- 895
- 896

897 Workflows and decision making:

- 898 • I'm trying to understand the details of how taxonomists do their work and the decisions they make along the
- 899 way. Could you talk me through your workflow?
- 900
- 901 • Could you draw me a schematic of your workflow, i.e. a workflow diagram?
- 902 • How do you standardize your descriptions across species in a single work?
- 903 • How do you standardize your descriptions between works?
- 904

905 Data collection, management and creation:

- 906 • How do store your observations? In what format?
- 907 • How do you make your observations available to others?
- 908 • At what point in your workflow do you record anatomical descriptions
- 909 • When do you begin to describe a taxon?
- 910 • How do you know when it is time to describe species
- 911 • How much time does it take to describe a species
- 912 • What is your goal when you describe a species?
- 913 • How are you "sure" that someone else understands what you do? What do you describe?
- 914 • How do you deal with visualizing your description'
- 915
- 916
- 917

918 Concluding questions

- 919 • Some people have said to us that they see taxonomic work as a process of standardizing. What do you think?
- 920 • What frustrates you about doing taxonomy?
- 921 • Is there a particular "technology" that you know others have that you wish you had? That nobody has but you
- 922 wish existed?
- 923 • Do you know of anyone else that might be willing to be interviewed?
- 924
- 925

## 926 **B HACKATHON WORKBOOK TEMPLATE**

927 **Title**

928 Initial contributors:

929 *Note: Consider including words that might classify the use-case*

930 **Background & Goals**

- 931 • For the purposes of this hackathon is this a "big", "medium", or "little" use case

932 Manuscript submitted to ACM

- Very brief, a short story, three or four sentences that lets both domains (biologists/developers) get a feel for the problem;
- A statement or description of where it fits in the design space (e.g. this is about manipulating matrices?) Why do you want to do this?

### **Workflow/Approach (Relevant Use Cases)**

- A bullet list and/or (UML-style) diagram that itemizes both human and computational steps in data collection, processing, and analysis
- Each step should seek to address the following - 1) input, 2) process, 3) output

### **Constraints**

- A list of challenges, with a focus on the technical issues (e.g. must interoperate with a certain database)
- May include definition of core concepts (e.g. must be applicable to specimen level data; must allow for quantitative states)

### **Implications for Design**

- Think of this as a brainstorming phase: how might we design tools to achieve our research goals (within the above constraints)?
- What do you want? what would you like?
- What don’t you like about what you have to do currently?
- Computers are really good at remembering, doing boring repetitive things, checking, counting, keeping track of things, etc – where do you have to do things like that?
- Are there new technological opportunities that we might take advantage of (e.g. using ontologies, using microphotography, other tools)?
- Detail functional requirements (define what must minimally happen)

### **Proposed widgets/interfaces**

- Must include an illustration/wireframe
- May reference specific implications for design
- Should identify preconditions for using this widget (e.g. a hierarchy of taxonomic names is available)

### **Summary for software developer**

Is done last. Consider- in formulating this use case where did you spend most of your time? What’s surprising about this use case? What’s a priority derived from this use case? What’s counter-intuitive to the assumptions a developer may have?

### **ACKNOWLEDGMENTS**

The project was funded by the National Science Foundation under grant number ABI-1356515: “Collaborative Research: ABI Innovation: Rapid prototyping of semantic enhancements to biodiversity informatics platforms.” For further information about this project, please visit <https://bioip.github.io/>. We are immensely grateful to all our participants throughout this project, particularly those who hosted us at their institutions, and participated in our Hackathon. We

989 thank István Mikó and his team for the conceptual discussions and insights that arose from his practical application  
 990 of the ideas here to newly published species descriptions. We thank Chris Nixon and Jinlong “Kenny” Guo for their  
 991 contributions to data collection and early data analysis. Thanks to Katrina Fenlon for her insightful comments regarding  
 992 the varied meanings of the word “articulation.” Finally, thank you to our reviewers and the program committee for their  
 993 patience and incredibly helpful reviews; their comments improved this manuscript immeasurably and we appreciate  
 994 their hard work.  
 995  
 996  
 997

## 998 REFERENCES

- 999 2018. "Articulation". (2018). <https://www.merriam-webster.com/dictionary/articulation>
- 1000 2018. Taxonomy. <https://en.wikipedia.org/w/index.php?title=Taxonomy&oldid=848162523> Page Version ID: 848162523.
- 1001 Mark S. Ackerman, Juri Dachtera, Volkmar Pipek, and Volker Wulf. 2013. Sharing Knowledge and Expertise: The CSCW View of Knowledge Management. *Computer Supported Cooperative Work (CSCW)* 22, 4 (2013), 531–573. <https://doi.org/10.1007/s10606-013-9192-8>
- 1002 Hanne Albrechtsen and Elin K. Jacob. 1998. The Dynamics of Classification Systems as Boundary Objects for Cooperation in the Electronic Library. *Library Trends* 47, 2 (1998), 293–312.
- 1003 Hanne Albrechtsen and Annelise Pejtersen. 2003. Cognitive Work Analysis and Work Centered Design of Classification Schemes. *International Journal of Knowledge Organisation* 30 (2003), 213–227.
- 1004 Wayne P Armstrong. 1999. Leaf Terminology. <http://www2.palomar.edu/users/warmstrong/termf2.htm>. Accessed: 2018-4-17.
- 1005 Esa Auramäki, Mike Robinson, Ann Aaltonen, Mikko Kovalainen, Arja Liinamaa, and Taina Tuuna-Väiskä. 1996. Paperwork at 78kph. In *Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work*. ACM, 370–379.
- 1006 James P. Balhoff, István Mikó, Matthew J. Yoder, Patricia L. Mullins, and Andrew R. Deans. 2013. A Semantic Model for Species Description Applied to the Ensign Wasps (Hymenoptera: Evaniidae) of New Caledonia. *Systematic Biology* 62, 5 (2013), 639–659. <https://doi.org/10.1093/sysbio/syt028>
- 1007 Liam J Bannon and Kjeld Schmidt. 1989. CSCW - Four Characters in Search of a Context. *DAIMI Report Series* 18, 289 (1989).
- 1008 Matthew J. Bietz and Charlotte P. Lee. 2012. Adapting cyberinfrastructure to new science: tensions and strategies. ACM Press, 183–190. <https://doi.org/10.1145/2132176.2132200>
- 1009 Jeanette Blomberg and Helena Karasti. 2012. Positioning ethnography within Participatory Design. *Routledge Handbook of Participatory Design* September (2012), 86–116.
- 1010 Geoffrey C Bowker. 1999. The game of the name: Nomenclatural instability in the history of botanical informatics. *Science Information Systems* (1999), 74.
- 1011 Geoffrey C Bowker. 2000. Biodiversity Datadiversity. *Soc. Stud. Sci.* 30, 5 (2000), 643–683.
- 1012 Geoffrey C Bowker and Susan Leigh Star. 2000. *Sorting Things Out: Classification and Its Consequences*. MIT Press.
- 1013 Andre Breitenfeld, Maximilian Mackeprang, and Ming-Tung Hong. 2018. Enabling Structured Data Generation by Nontechnical Experts. *Mensch und Computer 2017-Tagungsband: Spielend einfach interagieren* 17 (2018), 181.
- 1014 Julia Bullard. 2016. Warrant as a means to study classification system design. *Journal of Documentation* 73, 1 (2016), 75–90. <https://doi.org/10.1108/JD-06-2016-0074>
- 1015 Kathy Charmaz. 2014. *Constructing grounded theory* (2nd edition ed.). Sage, London ; Thousand Oaks, Calif. OCLC: ocn878133162.
- 1016 George Chin, Jr. and Carina S. Lansing. 2004. Capturing and Supporting Contexts for Scientific Data Sharing via the Biological Sciences Collaboratory. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work (CSCW '04)*. ACM, 409–418. <https://doi.org/10.1145/1031607.1031677>
- 1017 The Library of Congress. [n. d.]. LC Linked Data Service: Authorities and Vocabularies (Library of Congress). <http://id.loc.gov/ontologies/lcc.html>
- 1018 Michael Cook. 2002. Periodic Table in OWL. [http://explore.dublincore.net/learning\\_resource/periodic-table-in-owl/](http://explore.dublincore.net/learning_resource/periodic-table-in-owl/)
- 1019 Mark J. Costello, Simon Wilson, and Brett Houlding. 2013. More Taxonomists Describing Significantly Fewer Species per Unit Effort May Indicate That Most Species Have Been Discovered. 62 (2013).
- 1020 Hong Cui. 2010. Semantic annotation of morphological descriptions: an overall strategy. *BMC Bioinformatics* 11, 1 (2010), 278.
- 1021 Hong Cui. 2012. CharaParser for fine-grained semantic annotation of organism morphological descriptions. *Journal of the American Society for Information Science and Technology* 63, 4 (2012), 738–754. <https://doi.org/10.1002/asi.22618>
- 1022 Cristina David, Pascal Kesseli, and Daniel Kroening. 2017. Kayak: Safe Semantic Refactoring to Java Streams. *arXiv:1712.07388 [cs]* (2017). [arXiv:1712.07388](http://arxiv.org/abs/1712.07388)
- 1023 Andrew R Deans, Matthew J. Yoder, and James P. Balhoff. 2012. Time to change how we describe biodiversity. *Trends in Ecology & Evolution* 27, 2 (2012), 78–84.
- 1024 Ixchel M. Faniel and Trond E. Jacobsen. 2010. Reusing Scientific Data: How Earthquake Engineering Researchers Assess the Reusability of Colleagues' Data. *Computer Supported Cooperative Work (CSCW)* 19, 3 (2010), 355–375. <https://doi.org/10.1007/s10606-010-9117-8>
- 1025 Nico M. Franz, Mingmin Chen, Parisa Kianmajd, Shizhuo Yu, Shawn Bowers, Alan S. Weakley, and Bertram LudÅdscher. 2016. Names are not good enough: Reasoning over taxonomic change in the *Andropogon* complex 1. *Semantic Web Interoperability, Usability, Applicability* 7, 6 (2016), 645–667. <https://doi.org/10.3233/SW-160220>
- 1026  
 1027  
 1028  
 1029  
 1030  
 1031  
 1032  
 1033  
 1034  
 1035  
 1036  
 1037  
 1038  
 1039  
 1040

Manuscript submitted to ACM



1041  
1042  
1043  
1044  
1045  
1046  
1047  
1048  
1049  
1050  
1051  
1052  
1053  
1054  
1055  
1056  
1057  
1058  
1059  
1060  
1061  
1062  
1063  
1064  
1065  
1066  
1067  
1068  
1069  
1070  
1071  
1072  
1073  
1074  
1075  
1076  
1077  
1078  
1079  
1080  
1081  
1082  
1083  
1084  
1085  
1086  
1087  
1088  
1089  
1090  
1091  
1092

Elihu M. Gerson and Susan Leigh Star. 1986. Analyzing Due Process in the Workplace. *ACM Trans. Inf. Syst.* 4, 3 (July 1986), 257–270. <https://doi.org/10.1145/214427.214431>

Barney G. Glaser and Anselm L. Strauss. 2017. *Discovery of Grounded Theory: Strategies for Qualitative Research* (1st electronic edition ed.). Routledge, New York.

Jane Greenberg, Angela Murillo, John Kunze, Sarah Callaghan, Rob Guralnick, Nassib Nassar, Karthik Ram, Greg JanÅle, and Christopher Patton. 2013. Metadictionary: Advocating for a community-driven metadata vocabulary application.

Nicola Guarino and Christopher A. Welty. 2004. An Overview of OntoClean. In *Handbook on Ontologies*. Springer Berlin Heidelberg, 151–171. [https://doi.org/10.1007/978-3-540-24750-0\\_8](https://doi.org/10.1007/978-3-540-24750-0_8)

Patricia Harpring. 2010. *Introduction to controlled vocabularies: terminology for art, architecture, and other cultural works*. <http://www.getty.edu/publications/virtuallibrary/160606018X.html> OCLC: 897788032.

Christine Hine. 2006. Databases as Scientific Instruments and Their Role in the Ordering of Scientific Work. *Social Studies of Science* 36, 2 (apr 2006), 269–298. <https://doi.org/10.1177/0306312706054047>

Christine Hine. 2008. *Systematics as cyberscience: computers, change, and continuity in science*. The MIT Press.

Robert Hoehndorf, Paul N. Schofield, and Georgios V. Gkoutos. 2015. The role of ontologies in biological and biomedical research: a functional perspective. *Briefings in Bioinformatics* (2015), bbv011. <https://doi.org/10.1093/bib/bbv011>

Morgan D. Jackson. 2015. Gender Issues in Taxonomy: more than just Latin. <http://www.biodiversityinfocus.com/blog/2015/05/07/gender-issues-in-taxonomy-more-than-just-latin/>

Eva Jansen. 2013. A Review of Boundary Objects in Classification Research. *Advances in Classification Research Online* 23, 1 (2013), 52.

Marina Jirotko, Charlotte P Lee, and Gary M Olson. 2013. Supporting Scientific Collaboration: Methods, Tools and Concepts. *Computer Supported Cooperative Work (CSCW)* 22, 4-6 (2013), 667–715.

Lucas N. Joppa, David L Roberts, and Stuart L Pimm. 2011. The population ecology and social behaviour of taxonomists. *Trends in Ecology & Evolution* 26, 11 (2011), 551–553.

Helena Karasti, Karen S Baker, and Eija Halkola. 2006. Enriching the notion of data curation in e-science: data managing and information infrastructuring in the long term ecological research (LTER) network. *Computer Supported Cooperative Work (CSCW)* 15, 4 (2006), 321–358. <https://doi.org/10.1007/s10606-006-9023-2>

Helena Karasti, Karen S. Baker, and Florence Millerand. 2010. Infrastructure Time: Long-term Matters in Collaborative Development. *Computer Supported Cooperative Work (CSCW)* 19, 3 (2010), 377–415. <https://doi.org/10.1007/s10606-010-9113-z>

Helena Karasti and Jeanette Blomberg. 2017. Studying Infrastructuring Ethnographically. *Comput. Support. Coop. Work* 27, 2 (2017), 233–265.

Toshiaki Katayama, Kazuharu Arakawa, Mitsuteru Nakao, Keiichiro Ono, Kiyoko F Aoki-Kinoshita, Yasunori Yamamoto, Atsuko Yamaguchi, Shuichi Kawashima, Hong-Woo Chun, Jan Aerts, Bruno Aranda, Lord Hendrix Barboza, Raoul Jp Bonnal, Richard Bruskwiech, Jan C Bryne, José M Fernández, Akira Funahashi, Paul Mk Gordon, Naohisa Goto, Andreas Groscurth, Alex Gutteridge, Richard Holland, Yoshinobu Kano, Edward a Kawas, Arnaud Kerhornou, Eri Kibukawa, Akira R Kinjo, Michael Kuhn, Hilmar Lapp, Heikki Lehvaslaiho, Hiroyuki Nakamura, Yasukazu Nakamura, Tatsuya Nishizawa, Chikashi Nobata, Tamotsu Noguchi, Thomas M Oinn, Shinobu Okamoto, Stuart Owen, Evangelos Pafilis, Matthew Pocock, Pjotr Prins, René Ranzinger, Florian Reisinger, Lukasz Salwinski, Mark Schreiber, Martin Senger, Yasumasa Shigemoto, Daron M Standley, Hideaki Sugawara, Toshiyuki Tashiro, Oswaldo Trelles, Rutger a Vos, Mark D Wilkinson, William York, Christian M Zmasek, Kiyoshi Asai, and Toshihisa Takagi. 2010. The DBCLS BioHackathon: standardization and interoperability for bioinformatics web services and workflows. The DBCLS BioHackathon Consortium\*. *Journal of biomedical semantics* 1, 1 (2010), 8. <https://doi.org/10.1186/2041-1480-1-8>

Matthias Korn, Marén Schorch, Volkmar Pipek, Matthew Bietz, Carsten Österlund, Rob Procter, David Ribes, and Robin Williams. 2017. E-Infrastructures for Research Collaboration: The Case of the Social Sciences and Humanities. In *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW ’17 Companion)*. ACM, 415–420. <https://doi.org/10.1145/3022198.3022655>

Tobias Kuhn, Christine Chichester, Michael Krauthammer, Nüría Queralt-Rosinach, Ruben Verborgh, George Giannakopoulos, Axel-Cyrille Ngonga Ngomo, Raffaele Vigiñanti, and Michel Dumontier. 2016. Decentralized provenance-aware publishing with nanopublications.

John Kunze, Scout Calvert, Jeremy D DeBarry, Matthew Hanlon, Greg Janée, and Sandra Sweat. 2017. Persistence Statements: Describing Digital Stickiness. *Data Science Journal* 16, 0 (Aug. 2017), 39.

Hilmar Lapp, Sendu Bala, James P. Balhoff, Amy Bouck, Naohisa Goto, Mark Holder, Richard Holland, Alisha Holloway, Toshiaki Katayama, Paul O. Lewis, Aaron J. Mackey, Brian I. Osborne, William H. Piel, Sergei L. Kosakovsky Pond, Art F.Y. Poon, Wei-Gang Qiu, Jason E. Stajich, Arlin Stoltzfus, Tobias Thierer, Albert J. Vilella, Rutger a. Vos, Christian M. Zmasek, Derrick J. Zwickl, and Todd J. Vision. 2007. The 2006 NESCent Phyloinformatics Hackathon: A Field Report. *Evolutionary Bioinformatics* 3 (2007), 287–296.

Jane Li, Toni Robertson, and Christian MÅijller-Tomfelde. 2012. Distributed Scientific Group Collaboration Across Biocontainment Barriers. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (CSCW ’12)*. ACM, 1247–1256. <https://doi.org/10.1145/2145204.2145392>

Robert M. Losee. 2006. Optimal User-Centered Knowledge Organization and Classification Systems: Using Non-reflected Gray Codes. *Journal of Digital Information* 2, 3 (2006). <https://journals.tdl.org/jodi/index.php/jodi/article/view/47>

David R. Millen. 2000. Rapid ethnography: time deepening strategies for HCI field research. *Proceedings of the 3rd Conference on Designing Interactive Systems: processes, practices, methods, and techniques (DIS ’00)* (2000), 280–286. <http://dl.acm.org/citation.cfm?id=347763>

Barend Mons and Jan Velterop. 2009. Nano-Publication in the e-science era. In *Workshop on Semantic Web Applications in Scientific Discourse (SWASD 2009)*. 14–15.

- 1093 Robert D. Montoya and Seth Erickson. 2017. Anachronism in Global Information Systems: the cases of Catalogue of Life and Unicode. In *iConference*  
1094 *Proceedings*, Vol. 2. 123–128. <https://doi.org/10.9776/17110>
- 1095 John S Noyes. 1988. Encyrtidae (Insecta: Hymenoptera). *Fauna of New Zealand* 13 (1988).
- 1096 Gary M. Olson and Judith S. Olson. 2000. Distance Matters. *Human-Computer Interaction* 15, 2 (2000), 139–178. [https://doi.org/10.1207/S15327051HCI1523\\_4](https://doi.org/10.1207/S15327051HCI1523_4)
- 1097 Irene V. Pasquetto, Ashley E. Sands, Peter T. Darch, and Christine L. Borgman. 2016. Open Data in Scientific Settings: From Policy to Practice. ACM Press, 1585–1596. <https://doi.org/10.1145/2858036.2858543>
- 1098 Derek J. De Solla Price. 1965. Networks of Scientific Papers. *Science* 149, 3683 (1965), 510–515. <http://www.jstor.org/stable/1716232>
- 1100 Gorka Puente and Oscar Diaz. 2012. Wiki Refactoring as Mind Map Reshaping. In *Lecture Notes in Computer Science*. 646–661.
- 1101 Dave Randall, Rob Procter, Yuwei Lin, Meik Poschen, Wes Sharrock, and Robert Stevens. 2011. Distributed ontology building as practical work. *Int. J. Hum. Comput. Stud.* 69, 4 (2011), 220–233.
- 1102 David P. Randall, E. Ilana Diamant, and Charlotte P. Lee. 2015. Creating Sustainable Cyberinfrastructures. ACM Press, 1759–1768. <https://doi.org/10.1145/2702123.2702216>
- 1103 Madhu C. Reddy, Paul Dourish, and Wanda Pratt. 2006. Temporality in Medical Work: Time also Matters. *Computer Supported Cooperative Work (CSCW)* 15, 1 (jan 2006), 29–53. <https://doi.org/10.1007/s10606-005-9010-z>
- 1104 David Ribes. 2014. The kernel of a research infrastructure. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. ACM, 574–587.
- 1105 David Ribes and Thomas Finholt. 2009. "The Long Now of Technology Infrastructure: Articulating Tensions in Development. *Journal of the Association for Information Systems* 10, 5 (2009). <http://aisel.aisnet.org/jais/vol10/iss5/5/>
- 1106 Dana Rotman, Prociata Kezia, Derek Hansen, Cynthia Sims Parr, and Jennifer Preece. 2012. Supporting content curation communities: The case of the Encyclopedia of Life. *Journal of the American Society for Information Science and Technology* 63, 6 (2012), 1092–1107. <https://doi.org/10.1002/asi.22633>
- 1107 Andrew Russell and Lee Vinsel. 2016. Hail the maintainers. *Aeon Essays* (2016). <https://aeon.co/essays/innovation-is-overvalued-maintenance-often-matters-more>
- 1108 Cristina Sarasua, Elena Simperl, Natasha F Noy, Abraham Bernstein, and Jan Marco Leimeister. 2015. Crowdsourcing and the Semantic Web: A Research Manifesto. *Human Computation* 2, 1 (Aug. 2015).
- 1109 Aleksandra Sarcevic and Randall S Burd. 2009. Information handover in time-critical work. In *Proceedings of the ACM 2009 international conference on Supporting group work*. ACM, 301–310.
- 1110 Kjeld Schmidt. 2002. The Problem with 'Awareness': Introductory Remarks on 'Awareness in CSCW'. *Computer Supported Cooperative Work (CSCW)* 11, 3 (2002), 285–298. <https://doi.org/10.1023/A:1021272909573>
- 1111 Kjeld Schmidt and Liam Bannon. 2013. Constructing CSCW: The First Quarter Century. *Computer Supported Cooperative Work (CSCW)* 22, 4 (2013), 345–372. <https://doi.org/10.1007/s10606-013-9193-7>
- 1112 Kjeld Schmidt and Ina Wagner. 2004. Ordering systems: Coordinative practices and artifacts in architectural design and planning. *Computer Supported Cooperative Work (CSCW)* 13, 5-6 (2004), 349–408. <https://doi.org/10.1007/s10606-004-5059-3>
- 1113 Diane H. Sonnenwald. 2003. Expectations for a Scientific Collaboratory: A Case Study. In *Proceedings of the 2003 International ACM SIGGROUP Conference on Supporting Group Work (GROUP '03)*. ACM, 68–74. <https://doi.org/10.1145/958160.958171>
- 1114 Susan L. Star and James R. Griesemer. 1989. Institutional Ecology Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science* 19, 3 (aug 1989), 387–420. <https://doi.org/10.1177/030631289019003001>
- 1115 Susan L. Star and Karen Ruhleder. 1996. Steps toward an ecology of infrastructure: Design and access for large information spaces. *Information systems research* 7, 1 (1996), 111–134.
- 1116 Stephanie B. Steinhardt and Steven J. Jackson. 2015. Anticipation Work: Cultivating Vision in Collective Practice. ACM Press, 443–453. <https://doi.org/10.1145/2675133.2675298>
- 1117 Isabelle Stengers. 2018. *Another science is possible: A manifesto for slow science*. John Wiley & Sons.
- 1118 Anselm Strauss. 1985. Work and the Division of Labor. *The Sociological Quarterly* 26, 1 (1985), 1–19. <http://www.jstor.org/stable/4106172>
- 1119 Andrea K. Thomer, Michael B. Twidale, Jinlong Guo, and Matthew J. Yoder. 2016. Co-designing Scientific Software: Hackathons for Participatory Interface Design. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16)*. ACM, New York, NY, USA, 3219–3226.
- 1120 Gaurav Girish Vaidya. 2017. *Taxonomic Checklists as Biodiversity Data: How Series of Checklists can Provide Information on Synonymy, Circumscription Change and Taxonomic Discovery*. Ph.D. Dissertation. University of Colorado at Boulder, Ann Arbor, United States.
- 1121 Theresa Velden. 2013. Explaining Field Differences in Openness and Sharing in Scientific Communities. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work (CSCW '13)*. ACM, 445–458. <https://doi.org/10.1145/2441776.2441827>
- 1122 Janet Vertesi and Paul Dourish. 2011. The Value of Data : Considering the Context of Production in Data Economies. *CSCW 2011* (2011), 533–542.
- 1123 Markel Vigo, Caroline Jay, and Robert Stevens. 2014. Design insights for the next wave ontology authoring tools. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*.
- 1124 Markel Vigo, Caroline Jay, and Robert Stevens. 2015. Constructing Conceptual Knowledge Artefacts. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*.
- 1125 Manuscript submitted to ACM

1145  
1146  
1147  
1148  
1149  
1150  
1151  
1152  
1153  
1154  
1155  
1156  
1157  
1158  
1159  
1160  
1161  
1162  
1163  
1164  
1165  
1166  
1167  
1168  
1169  
1170  
1171  
1172  
1173  
1174  
1175  
1176  
1177  
1178  
1179  
1180  
1181  
1182  
1183  
1184  
1185  
1186  
1187  
1188  
1189  
1190  
1191  
1192  
1193  
1194  
1195  
1196

Amy Volda, Ellie Harmon, and Ban Al-Ani. 2011. Homebrew databases: Complexities of everyday information management in nonprofit organizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 915–924.

Paul Warren, Paul Mulholland, Trevor Collins, and Enrico Motta. 2014. Using Ontologies. In *Knowledge Engineering and Knowledge Management (Lecture Notes in Computer Science)*. Springer, Cham, 579–590. [https://doi.org/10.1007/978-3-319-13704-9\\_43](https://doi.org/10.1007/978-3-319-13704-9_43)

Matthew J. Yoder, İstván Miklós, Katja C. Seltmann, Matthew A. Bertone, and Andrew R. Deans. 2010. A Gross Anatomy Ontology for Hymenoptera. *PLoS ONE* 5, 12 (2010), e15991. <https://doi.org/10.1371/journal.pone.0015991>

Matthew J. Yoder, Michael B. Twidale, Andrea K. Thomer, Lars Vogt, Nico M. Franz, Jinlong Guo, Andrew R. Deans, and James Balhoff. 2018. Taxonomy and the production of semantic phenotypes. In *Application of Semantic Technologies in Biodiversity Science.*, Anne Thessen (Ed.). IOS Press/AKA Verlag.