

# Steps Towards an Infrastructure for Scholarly Synthesis

JOEL CHAN, University of Maryland

MATTHEW AKAMATSU, University of Washington

DAVID VARGAS, Vellum AI

LUKAS KAWERAU, Empiria Insight

MICHAEL GARTNER, Cybrarian Services

Sharing, reusing, and synthesizing knowledge is central to the research process, both individually, and with others. These core functions are not supported by our formal scholarly publishing infrastructure: instead of the smooth functioning of functional infrastructure, researchers resort to laborious "hacks" and workarounds to "mine" publications for what they need, and struggle to efficiently share the resulting information with others. Information scientists have proposed an alternative infrastructure based on the more appropriately granular model of a discourse graph of claims, and evidence, along with key rhetorical relationships between them. However, despite significant technical progress on standards and platforms, the predominant infrastructure remains steadfastly document-based. Drawing from infrastructure studies, we locate the current infrastructural bottlenecks in the lack of local systems that integrate discourse-centric models to augment synthesis work, from which an infrastructure for synthesis can be grown. Through 3 years of research through design and field deployment in a distributed community of hypertext notebook users, we elaborate a design vision of what can and should be built in order to grow a discourse-centric synthesis infrastructure: a thriving "installed base" of researchers authoring local, shareable discourse graphs to improve synthesis work, enhance primary research and research training, and augment collaborative research. We discuss how this design vision — and our empirical work — contributes steps towards a new infrastructure for synthesis, and increases HCI's capacity to advance collective intelligence and solve infrastructure-level problems.

CCS Concepts: • **Information systems** → **Digital libraries and archives**; • **Human-centered computing** → **Hypertext / hypermedia**; **Field studies**; **Empirical studies in HCI**.

Additional Key Words and Phrases: scholarly synthesis, hypertext, digital libraries, infrastructure

## ACM Reference Format:

Joel Chan, Matthew Akamatsu, David Vargas, Lukas Kawerau, and Michael Gartner. 2018. Steps Towards an Infrastructure for Scholarly Synthesis. 1, 1 (July 2018), 56 pages. <https://doi.org/10.1145/1122445.1122456>

## 1 INTRODUCTION

To advance the state of knowledge, researchers must *synthesize* what is currently known and unknown about problems. Effective synthesis generates new knowledge, integrating relevant theories, concepts, claims, and evidence into novel conceptual wholes [10, 105]. Synthesis may be supported by and manifested in various forms (e.g., a theory, a systematic

---

Authors' addresses: Joel Chan, [joelchan@umd.edu](mailto:joelchan@umd.edu), University of Maryland; Matthew Akamatsu, [akamatm@uw.edu](mailto:akamatm@uw.edu), University of Washington; David Vargas, [dvargas92495@gmail.com](mailto:dvargas92495@gmail.com), Vellum AI; Lukas Kawerau, [lukas@kawerau.org](mailto:lukas@kawerau.org), Empiria Insight; Michael Gartner, [michael@cybrarian.net](mailto:michael@cybrarian.net), Cybrarian Services.

---

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 ACM 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 Association for Computing Machinery.

Manuscript submitted to ACM

Manuscript submitted to ACM

1

or integrative literature review, or a model of a design space). The advanced understanding from synthesis can be a powerful force multiplier for choosing effective studies and operationalizations [72, 91, 110]. Unfortunately, synthesis work is frequently experienced as arduous [37, 42, 56], documented as extremely cost-intensive [13, 37, 74, 81, 98], and frequently absent in published research outputs, such as literature reviews in doctoral dissertations [12, 49, 67] and published papers [3, 4, 9, 39, 73, 110]. **What, if anything, can be done to make effective synthesis more commonplace? And what solutions might HCI contribute to complement the predominant focus on institution- and incentive-level interventions from a meta-research perspective?**

Information scientists have offered a compelling diagnosis of the problem of synthesis: there is a fundamental mismatch between researchers' information needs for synthesis work, and the document-centric data model of our predominant scholarly communication infrastructure. To do effective synthesis work smoothly, researchers need ready access to information artifacts at the level of claims, theories or empirical results [10, 29, 99]. These information artifacts are needed to grapple with the key questions that constitute the work of synthesis. For example, what empirical *evidence* supports/opposes the major *claims* under consideration? What are some alternative *claims* that are also consistent with the major *evidence* for our key claims or hypotheses? How might we weave together systems of discursively related *claims* into a coherent theory that can explain the key phenomena we seek to understand? What are the respective *evidence* bases (and that support theory X vs. theory Y? And what key points of (in)consistency in the *evidence* base might point to areas of productive controversy [105]?

To meet these information needs, information scientists propose that researchers need an infrastructure based on a *discourse-oriented* data model for the representation of scientific knowledge [18, 24, 30, 59, 99]. One prominent example is the micropublication model [24], which models scientific knowledge as a **discourse graph** of claims, contextualized by relevant context, such as authorship, evidence, and methods.

However, despite their significant potential for accelerating synthesis, discourse graphs have not yet gained widespread acceptance [59]; instead, the predominant scholarly communication infrastructures remain steadfastly coarse-grained and document-based. For example, digital libraries and search engines such as Google Scholar, PubMed, and Semantic Scholar, primarily index topics and documents, not the claims, theories or results that researchers need for synthesis. Citation databases largely do not show *why* scholarly works cite each other: we cannot trace lines of evidence, or how theories are used and applied in different settings, or quickly assemble collections of consistent or contradictory claims or results: Scite.ai's distinction between supportive vs. contradictory citations, and Semantic Scholar's distinctions between "background", "methods", and "results" citations, take steps towards this deeper understanding, but are still operating on the paper as a fundamental unit.

Understanding the mismatch between researchers' information needs and the basic data model of our current scholarly communication infrastructure clarifies why synthesis work is hard. Lacking a true infrastructure for synthesis, researchers make do with an assemblage of workarounds and hacks to find and synthesize relevant literature: they cobble across published syntheses; piece together leads on papers and authors and verbal statements of key ideas from knowledgeable colleagues; and engage in laborious citation tracing, manual reference checks, and extensive error-prone keyword search and screening of abstracts and full-text. From this emerging collection, researchers then construct *ad-hoc* databases of claims and data, strewn across annotations in PDFs, notes and partial drafts in Google Docs and spreadsheets [15, 48, 83, 113]. This bespoke database and system may work well enough for individual projects, but is rarely systematically transferred across projects and people, even within the same research group, leaving further related projects or people to repeat the arduous process, or worse: settle for subpar synthesis, or focus on diminishing low-hanging fruit instead of carefully and creatively conceptualizing new research directions.

In this paper, we build on this work to frame the problem of synthesis as a problem of **infrastructure development**. We conceptualize *infrastructure* here in the specific sense from infrastructure studies [34]. We recognize that infrastructure is fundamentally *sociotechnical* [64]: it consists of technical components intertwined with, and sustained/constrained by, social phenomena and structures (e.g., culture, practices, institutions). We also recognize that infrastructures cannot be designed or constructed *a priori*, but instead, are **grown** and accreted over past infrastructures [1, 33, 104]. Edwards et al elaborate on this intuition by drawing three key implications from historical models of infrastructural development from history and sociology [51]: “First, true infrastructures only begin to form when locally constructed, centrally controlled *systems* are linked into *networks* and *internetworks* governed by distributed control and coordination processes. Second, infrastructure formation typically starts with *technology transfer* from one location or domain to another; adapting a system to new conditions introduces technical variations as well as social, cultural, organization, legal, and financial adjustment. Third, infrastructures are consolidated by means of *gateways* that permit the linking of heterogeneous systems into networks” (p. 8; emphasis in original).

Mapping this formulation of the infrastructure development problem to our setting, we see that existing work on the discourse graph data model has laid a foundation for *gateways* that can weave networks of networks into larger scale infrastructure for synthesis. The remaining problem is an HCI problem: how to develop *local systems* that enhance synthesis, but leave seeds for development into infrastructure: (socio)technical means of *networking* these systems in a distributed fashion (e.g., via peer-to-peer networking), and *technology transfer* across different contexts. This formulation of the remaining open problem aligns with researchers’ diagnosis of the current main blocker to progress towards a discourse-centric synthesis infrastructure: the lack of sociotechnical means — such as people to do the work of creating discourse graphs, and tools for them to do this work — for creating and sustaining this infrastructure [43, 59, 60, 84]. This paper therefore explores the following core research question: **What might it look like to integrate the seeds of a discourse-centric infrastructure into local scientific practices, and the collaborative and institutional structures of research?**

We approached this question using **Research through Design** (RTD) [118]: we *designed* and built a software extension for a hypertext notebook — used by thousands of researchers for notetaking, lab management, and writing — that could structure and share research knowledge in a discourse-centric data model. We then *deployed* our software prototype in this user community over the course of 2.5 years, iterated on the design in response to authentic usage. Finally, we *reflected* on conceptual and empirical findings from the deployment, including direct interactions and observations with 48 researchers in the user community.

Our research through design work resulted in concrete design patterns, prototypes, and empirical evidence of authentic usage that elaborated a design vision of what can and should be built — at least at the core/base — in order to grow a discourse-centric synthesis infrastructure: a thriving “installed base” [46, 104] of researchers authoring local discourse graphs to improve synthesis work (§5.1), enhance primary research and research training (§5.2), and augment collaborative research (§5.3). The work of discourse graph authoring is technically enabled by local systems of incrementally formalizable (§6.1 and locally extensible §6.2) discourse graph authoring mechanisms. These local systems can be transferred across a diverse ecosystem of personal and lab hypertext notebooks (§6.3), that can then be networked in a peer-to-peer fashion using the discourse graph model as a shared data exchange protocol.

This paper’s primary contribution is therefore **a design vision for — and concrete steps towards — a new infrastructure for synthesis**. larger infrastructure for effective scholarly synthesis. Because research synthesis shares many characteristics with — and interacts with — other forms of complex (and collective) creative work, such as

sensemaking (e.g., intelligence analysis), design innovation, and collective deliberation and governance, our research also increases HCI’s capacity to advance collective intelligence.

## 2 RELATED WORK

### 2.1 Knowledge infrastructures in science

Our focus on a discourse-centric infrastructure for synthesis here complements existing infrastructure studies on knowledge infrastructure, which tend to focus primarily on *data* (collection, sharing, and reuse) [14, 27, 32, 35, 87], not on discourse and synthesis of granular concepts and claims. For example, Edwards et al [32] described the large-scale distributed research infrastructure for climate studies and modeling, and Mosconi et al [76] described tensions between data sharing mandates and large-scale infrastructures on the one hand, and interdisciplinary researchers’ data management and sharing practices on the ground.

Additionally, since a key analytic strategy of infrastructural inversion looks for moments of maintenance, upgrade, and breakdown [86], much of this research identifies where infrastructure is *not* working [76]. While this perspective is valuable for understanding what *is*, it is limited for understanding what *could be* (our focus). Our work here aligns more with constructively oriented work like Bowker’s [17] synthesis of insights from science studies to propose design principles for constructing a global biodiversity database, and Young et al’s [117], ambitious design arc from an infrastructural investigation of land change scientists’ needs for synthesis.

### 2.2 New Conceptual Models for Synthesis Infrastructures

The proposal for a discourse-centric infrastructure that we explore here is situated in a larger array of related proposals — including ontologies, semantically rich data models, and other metadata and linked data standards — for new modes of knowledge representation, sharing, and transfer [30, 60, 85]. In this paper, we use the term **"discourse graph"** to refer to a suite of information models [18, 23, 24, 31, 44, 70, 99, 100] that share a common underlying model for representing scientific discourse: one that distills traditional forms of publication down into more granular, formalized knowledge *claims*, linked to supporting evidence and *context* through a network or *graph* model. Notable inspirations for this our work here include McCrickard’s claims model for representing design knowledge [70, 71]; the micropublications model of claims and evidence [24], SEPIO ontology for representing assertions and evidence lines [18]; and the nanopublications model [44] that aims to bridge less formal natural-language scientific assertions to more formalized machine-readable assertions [59] that can integrate into more sophisticated “semantic publishing” infrastructures [60].

We choose the term discourse graph to refer to this family of data models partly because of the conceptual roots of these information models in argumentation theory, such as Toulmin’s model of argumentation [107], and to distinguish them from ontologies and other concept-centric forms of knowledge graphs: with discourse graphs, the emphasis is on representing and relating knowledge claims (rather than concepts) as the central unit, and emphasizing linking and relating these claims (rather than categorizing or filing them).

To understand the potential of this data model as a foundation for an infrastructure for synthesis, consider a researcher who wants to understand what interventions might be most promising for mitigating online harassment. To synthesize a formulation of this complex interdisciplinary problem that can advance the state of the art, she needs material that can help her work through detailed answers to a range of questions. For example, which theories have the most empirical support in this particular setting? What are the key phenomena to keep in mind when designing an intervention?



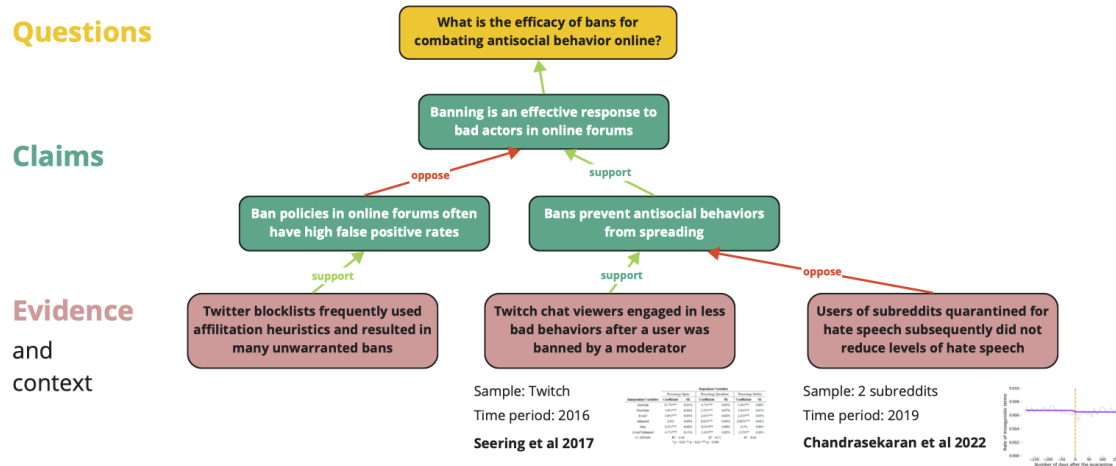


Fig. 1. Example discourse graph (with claims and associated context) for theories and findings on effects of bans on bad actors in online forums.

What intervention patterns (whether technical, behavioral, or institutional) have been proposed that are both a) judged on theoretical and circumstantial grounds as likely to be effective in this setting, and b) lacking in direct evidence for efficacy?

The answers to these questions cannot be found in the common representations of scientific knowledge in current tools, such as titles and abstracts of research papers, groupings of papers by area, or citation or authorship networks. Instead, the answers lie at the more granular level of theoretical and empirical **claims** or statements. For example, “viewers in a Twitch chat engaged in less bad behaviors after a user was banned by a moderator for bad behavior” [95], and “users of subreddits quarantined for hate speech subsequently did not reduce levels of hate speech” [21] are empirical claims that interrelate in complex ways, each supporting claims and theories that may be in tension with each other. Granular representations of claims is crucial not just for precisely *finding* relevant information to inform the synthesis, but also for *constructing* more appropriately complex arguments and theories, by connecting statements in logical and discursive relationships. Beyond operating at the claim level, our researcher will also need to work through a range of contextual, methodological details about empirical **evidence**. For example, to compare, judge, and develop claims (e.g., what has been established with sufficient certainty, where the frontier might be), she would need to know, for example, which empirical evidence came from which measures (e.g., self-report, behavioral measures), and the extent to which results have been replicated across a variety of settings (e.g., different research groups, years, platforms, scales).

Figure 1 illustrates a discourse graph that might support such an analysis. In this discourse graph, the primary unit of analysis of claims matches the information needs of synthesis. Contextual entities and information, such as methodological details and metadata, are explicitly attached to evidence in the discourse graph, enabling direct analysis of claims with their evidentiary context, and critical engagement, integration, or reinterpretation of individual findings. And discourse units in the discourse graph can have many-to-many relationships to support composition of more complex arguments and theories. If appropriately stable domain-specific vocabulary or ontologies are available, a set of

claims may also be modeled as defining a subgraph of a more granular knowledge or causal graph, supporting more sophisticated theoretical synthesis that is still connected to empirical evidence.

There have been attempts to integrate discourse graphs into scientific communities of practice: for example, the ScholOnto model [100] was supporting remote PhD mentoring and distributed collaborations; SWAN [23] was integrated into the successful Alzforum online research community for Alzheimers research [25]; and the micropublication model was integrated into the Domeo scientist network [24]. However, for a variety of reasons, these nascent infrastructures are no longer active, and the impact of these deployments has not been empirically evaluated. This may partly be due to changes in funding and infrastructure for these research software, leading to deprecation of technical infrastructure. Other efforts might not have made it past the experimental prototype stage for similar reasons (lack of funding, incentives). However, the pivot of some information models into more educational or general-purpose applications [66] is suggestive of the open problems around realizing a discourse-centric infrastructure in scientific practice. As we note above, there is some consensus that these problems center on the "human" side of the infrastructure such as tools and incentives for authors, and sociotechnical means of networking and sharing discourse graphs in practice [30, 60]. Our work joins and complements existing nascent explorations of different ways to foster sustainable contributions to synthesis infrastructures, such as controlled natural language wikis [58], markup integrations in word processors [45, 108], and annotation schemes in publication submission portals [19].

### 2.3 Existing and emerging tools and workflows for synthesis work

Our work is grounded in fundamental studies of the tools and workflows of scholars [80], which have yielded valuable descriptions of the information needs, pain points, and regularities (and differences) in scholarly workflows and tools. Examples include Ellis's classic study of the information-seeking patterns of academic researchers [36], O'Hara's diary study of PhD students' use of library documents and reading practices [79], Ribaupierre et al's description of scholars' strategic reading practices [29], and Blake and Pratt's detailed description and model of the information behaviors of scientists conducting systematic reviews [10]. On the practice side, surveys of tool usage by scientists suggest that document-centric workflows continue to dominate. For example, Bosman et al. [15] reported from a large-scale online survey of approximately 20k researchers worldwide that reference management tools like EndNote, Mendeley, and Zotero were the most frequently mentioned tools for managing and using literature. These large-scale findings are corroborated by more in-depth qualitative investigations of researcher practices, which generally find the predominance of these document-centric tools, as well as mainstream general-purpose software like Microsoft Word and Excel for note-taking [48, 83, 113]. One notable study by Sawyer and colleagues [90] frames the landscape of tools, workflows, and practices of synthesis as "pre-infrastructure". They note that these are marked by assemblage, fragmentation, and appropriation, and are not mature infrastructure.

Some of this descriptive work provides empirical validation of the hypothesized benefits of discourse graphs for augmenting synthesis: for example, Ribaupierre et al [29] found that scientists who tested a prototype tool that allowed them to search over a literature by rhetorical elements (e.g., findings, methods, definitions) self-reported higher signal-to-noise ratio in results when searching for specific findings (e.g., "show all findings of studies that have addressed the issue of gender equality in terms of salary"), compared to using a standard keyword search interface. There is also documentation of desire paths of scientists adopting niche tools with affordances that resemble discourse graphs. For example, there is a subculture of academic researchers who repurpose qualitative data analysis tools like NVivo and Atlas.ti to do literature reviews [6, 75, 102, 114]; it is notable that the key affordances of these tools — layers of contextualizability for excerpts (data) and themes (synthesis/claims) — are similar to those of discourse graphs. There is

also some adoption of niche specialized tools for literature sensemaking, such as LiquidText and Citavi, both of which support the composition of networks of claims that are directly linked to contextualizing excerpts from documents [75].

Because our infrastructure development strategy focuses on developing seeds for infrastructure by integrating into and augmenting local synthesis practices, the prototypes, design patterns, and empirical evidence in this paper also advance research on tools and workflows for synthesis work.

### 3 METHODOLOGY: RESEARCH THROUGH DESIGN

We explored our research questions using a Research through Design (RtD) methodology [118]. We chose RtD because the nature of our questions are about "the world that should be brought into being" [118, p. 178]: our questions cannot be descriptive, because there are no existing examples of a discourse-centric synthesis infrastructure. The "wicked" and messy nature of an infrastructure-level problem — with all of its complex sociotechnical entanglements and entrenched, competing interests — is also well-matched for the application of a designerly — vs. scientific — way of knowing [41, 118].

We followed Zimmerman and Forlizzi's [118] suggested 5-step cycle for Research through Design work: 1) *select* a design problem or opportunity, 2) conduct *design* work to develop an initial framing and prototypes, 3) *evaluate* the design prototype(s) through deployment with users, 4) *reflect* on the design and evaluation work and *disseminate* resulting insights, and 5) repeat. This paper reports analysis of insights from design work and artifacts and research data collected across the **Design** and **Evaluate** steps of our Research through Design work in this setting (see Fig. 2 for a visual summary). These phases of work spanned approximately 3 years, and involving direct interactions with 48 participants.

Our RtD approach was also conducted in a field-based and participatory manner. In particular, our work began as participatory in a participant observation sense: the first author is both researcher and participant/user in the community of tool users where we designed and deployed our tool. He drew on his own active participation and experiences to inform the design work and evaluation. Later, the work became participatory in a participatory design [89] and community-based research sense, as the second and fourth authors joined from the community to substantially develop the designs and insights and practice, fundamentally shaping the conceptual contributions of the work. We therefore drew on participatory methodology [111] — including reflexive, detailed field notes, regular research discussions with key community members, and co-authorship of the final manuscript — to ensure that community participation was rigorous and equitable.

#### 3.1 Setting

The general setting of this study is centered around a distributed community of researchers who have adopted general-purpose "hypertext notebooks" for research knowledge management. These notebooks share many features with wikis, with granular subdocuments and bi-directional hyperlinking between these subdocuments serving as a central organizing principle, as opposed to top-down hierarchies of folders and tags. There are also some interesting differences in affordances and values of these communities: for example, several tools, such as RoamResearch<sup>1</sup>, Logseq<sup>2</sup>, and Tana<sup>3</sup>, cross the hyperlinking wiki data structure with an "outliner" format, such that each bullet itself becomes a uniquely addressable subdocument that can be interlinked into knowledge and organizational structures; others,

<sup>1</sup><https://roamresearch.com/>

<sup>2</sup><https://logseq.com/>

<sup>3</sup><https://tana.inc>

<i>RtD Phase</i>	<i>DESIGN WORK &amp; ARTIFACTS</i>	<i>RESEARCH DATA COLLECTED</i>
2020		
Oct	<ul style="list-style-type: none"> <li>■ <u>Conceptual model</u> article detailing translation of base discourse model</li> </ul>	<ul style="list-style-type: none"> <li>■ <u>Observations</u> and <u>field notes</u> from participant observation</li> </ul>
<b>DESIGN</b> (~10 mo)	<ul style="list-style-type: none"> <li>■ <u>Video tours</u> of example implementation</li> <li>■ MVP <u>prototype</u> of Discourse Graph Extension in RoamResearch</li> </ul>	<ul style="list-style-type: none"> <li>■ <u>Design notes</u></li> </ul>
2021		
Aug	<ul style="list-style-type: none"> <li>■ <u>Enhancements</u> and new features in the Discourse Graph Extension prototype: initially in response to (and in collaboration with) users, then led by users</li> </ul>	<ul style="list-style-type: none"> <li>■ <u>Observations</u> and <u>field notes</u> from participant observation</li> <li>■ <u>Design notes</u></li> </ul>
<b>EVALUATE</b> (~30 mo)	<ul style="list-style-type: none"> <li>■ User-led <u>ports</u> of Discourse Graph workflow and extension to other tools</li> </ul>	<ul style="list-style-type: none"> <li>■ Usage <u>survey</u> deployed to users in September 2022 (<math>N=41</math>)</li> <li>■ <u>Data exports</u> from usage survey (<math>N=5</math>)</li> <li>■ Semi-structured <u>interviews</u> in September 2023 with users who indicated interest in a follow-up from the usage survey (<math>N=5</math>)</li> </ul>
2024		
Mar		<ul style="list-style-type: none"> <li>■ <u>Member checks</u> with key informants via email and calls in March 2024 (<math>N=7</math>)</li> </ul>

Fig. 2. Summary of design work and artifacts and research data collected across the Design and Evaluate steps of our Research through Design work.

such as Tinderbox<sup>4</sup>, embed the hyperlinking structure in a visual medium; and others, such as Obsidian<sup>5</sup>, heavily emphasize privacy and data ownership, overlaying wiki technology on plain text markdown files. From a scholarly primitives perspective [80], usage of these tools concentrates on the primitives of collecting, reading (including especially notetaking) and subsets of the writing process; frequently, traditional tools like word processors are still used for final drafting and typesetting stages of writing, but drawn heavily from the knowledge bases constructed in these tools.

There are no definitive numbers on how many researchers have adapted these general-purpose tools for synthesis work, but the following data points can give a sense of the size and boundaries of the community: as of the time of this writing, there were approximately 12,000 users on the Slack chat community for the RoamResearch tool, and approximately 800 users on a specialized Discord chat community for academic users of RoamResearch. We think the estimate of 5–10% of the total user base is a reasonable ballpark across the other tool communities, each of which have discussions and defined subcommunities for academic users, such as an academia channel in the Discord community of 8,000 users for the Obsidian tool, and 1500 users for the Logseq tool. Thus, this community of users is still nascent relative to mainstream tools like Google Docs and Microsoft Word, but it is substantial in size and active.

<sup>4</sup><http://www.eastgate.com/Tinderbox/>

<sup>5</sup><https://obsidian.md/>

### 3.2 Design Phase

Design work began in approximately December 2020 (and was mostly completed in summer of 2021), and centered around exploring how to instantiate the discourse graph model into a working system that enables scholars to write notes in as close to their normal formats as possible, but author shareable, machine-readable discourse graphs as a byproduct. The design work in this arc culminated in the initial working prototype of the RoamResearch Discourse Graph extension, which embodied a set of design hypotheses about how to integrate discourse graphing into everyday research workflows and enable downstream development of synthesis infrastructures. For brevity, we summarize the endpoint of this design work and resulting initial prototype in §4.

### 3.3 Evaluation Phase

Evaluation work began in approximately August 2021, and centered around testing our design hypotheses about the possibility and preconditions of 1) integrating discourse graphs into scholarly workflows, and 2) growing these "grassroots" discourse graphs into a new infrastructure for scholarly synthesis. Data collection has continued to the present, through sustained, serious usage of the RoamResearch Discourse Graph extension, and community engagement around the software prototype. Thus, the Deployment phase overlapped with the Reflection phase.

To test our design hypotheses, we released the extension in a number of different channels beginning the Fall of 2021. The first distribution channel was a set of three **notetaking courses** for academic users of Roam, one of which was led by LK, the fourth author. These courses are all online courses for users of Roam, and are led by lead users in the Roam community. Each lead user was an advanced PhD student at the time of the deployment. This integration came about because, as part of our participatory deployment and design strategy, the first author had solicited early feedback from each of these users on a call in August of 2021 (once an MVP of the prototype had been developed), and each of them expressed immediate strong interest in integrating this into their courses. The second distribution channel was the **wider Roam community**, which congregated on Twitter, blogs, Youtube, and a shared Slack channel.

At the time of this writing, the extension remains in active use. In this paper, however, we report our reflections on the following sources of data collected over the course of 30 months, from August 2021 to March 2024: 1) observations and field notes from participant observation, 2) design notes from design work, 3) a formal usage survey, 4) semi-structured interviews, 5) member checks with key informants, and 6) data exports from participants' lab notebooks.

*3.3.1 Observations and field notes.* The first author participated in wider discussions of workflows and tooling on Twitter, Slack, as well as synchronous Zoom calls. Structured direct observations were also conducted in a Discord channel dedicated to the extension in a Discord server for academic users of Roam. Finally, the first author held bi-weekly "open office hours" over Zoom with users of the extension. Field notes were recorded throughout these observations, including links to messages on the social media or chat threads, where appropriate/available, for later analysis.

*3.3.2 Design notes from design work.* All design and development discussions with the DV, the third author, were documented in a shared Roam notebook.

*3.3.3 Formal usage survey.* We distributed a usage survey in the Fall of 2022 to get a quantitative picture of usage. Key questions included demographics (e.g., occupation, location, research area, profession, source of learning about the extension), current usage patterns (e.g., current user vs. not, motivations for using/exploring the extension, which features of the extension are core/occasional/unused/unknown). The full survey items are shown in Appendix A.

The survey was distributed via a one-time popup in the extension, which users would see upon refresh of their Roam application, and also distributed over Twitter and in the Discord channel for users of the extension. We received  $N = 41$  total responses to the survey.

**3.3.4 Semi-structured interviews.** The first author conducted semi-structured interviews, each lasting approximately 60–90 minutes, with  $N = 5$  users who indicated an interest in a follow-up from the usage survey. The interview focused on whether/how they were still using the extension (and whether they'd be willing to share a csv export and/or screenshots of your discourse graph usage), impacts (positive or negative) the extension has had on their work since the last usage survey, changes in your synthesis practice (with or without the extension), and wishlist items for the extension (for ongoing development work). These interviews were recorded and transcribed.

**3.3.5 Member checks with key informants.** After initial analysis was completed in March 2024, the first author conducted member checks with  $N = 7$  key informants via email and calls.

**3.3.6 Data exports.** The usage survey requested an optional data export of the user's discourse graph. We were able to obtain user-provided exports of data from  $N = 4$  users (5, including the author's lab) from the usage survey approximately 12 months into the deployment (in Fall of 2022). These were neo4j-compatible csv exports of the nodes and edges.

This research study was approved by the University of Maryland's institutional review board.

### 3.4 Deployment Participants.

Since the software extension was released publicly into the community, and we (for privacy reasons) chose not to track specific usage on our servers from participants' machines, we do not have a precise quantitative picture of usage. We can, however, triangulate from at least two sources to give a quantitative sense of the userbase that is informing the insights from this arc. First, because of the way that Roam works with extensions, each time a Roam user reloaded their application, all software extensions would be (re)-downloaded and unpacked. This is not a precise measure of usage/coverage, since extensions that were left installed but were not in active use would still be downloaded at each reload of the application. Thus, the average daily downloads gave us a rough upper bound on the coverage of the extension. Based on the 2nd author's experience with other software extensions for Roam, an approximate estimate of active users given download counts is to divide the number of downloads by 10. As of November 2022, we had 320 daily downloads of the extension, so we estimate that the extension had about 30 daily active users as of November 2022. Second, 23 of the 41 ( 56%) respondents to our usage survey self-reported as current users of the extension, while 9 ( 22%) self-reported as "still exploring" the extension to assess fit, and 9 ( 22%) self-reported as not currently using the extension. Putting these two estimates together, we estimate our total participant base to be on the order of 30 daily active users.

Based on our usage survey, the majority (14/23, or 60%) of current users were academic researchers of some variety, with the remaining (9/23, or 40%) of users engaged in consulting, librarianship, medical training, independent research, and venture-building. Users hail from all over the world, including North America, China, Thailand, Germany, Chile, Austria, Spain, and New Zealand.

Also, in our usage survey, 11 respondents explicitly mentioned one of the three notetaking courses that integrated the extension, and 12 mentioned a mix of blogs, social media, or word-of-mouth as the first places they learned about the extension (the remaining 18 respondents did not describe how they learned about the extension).

### 3.5 Positionality and Rapport

The overall research design and conceptualization was initially led by the first author, an HCI researcher who is also a core user of the tool we extended. He participated as a “power user” in workflow and tool exploration, including the first “Roam tour”, published in April 2020, where he explored how to implement the “zetteltkasten method” in this new hypertext notebook medium. In general, his reputation as a “power user” in the hypertext notebook user community was at least as salient as his reputation as an HCI researcher. This gave him access and credibility he would not otherwise have gotten as outsiders, including partnerships with other community leaders who developed learning resources for the community of users. These partnerships were the first initial vector of the deployment of our design prototype. The first author’s additional identity as an HCI researcher was made plain to the community throughout the deployment, prominently advertised in primary announcements and resource sharing posts with the community, as well as in 1-on-1 and group calls. This is important because we obtained a waiver for documentation of informed consent, due to the distributed and often transient nature of interactions in the community.

Formative work on design requirements were shaped in part through extensive interactions with the second author — identified in the manuscript as MA — as fellow members of the user community: across several calls and interactions on the community forums. MA later led the extension of the discourse graph workflow and system to improve and transform primary research in his cell biology lab (see §5.2). The design and prototyping work for the software extension (see §4) was done in close collaboration with the third author, who is a key developer of software extensions for the tool we extended.

As the field study became increasingly participatory, two additional members of the community began to make increasingly integral contributions to the conceptualization and development of the key ideas in this paper, and have therefore been named as co-authors of this work (in keeping with norms of participatory research). The fourth author — identified in the manuscript as LK — saw the potential of the discourse graph workflow for enabling the learning of synthesis, and became a data collection partner, deploying the discourse graph software prototype we developed in and from the notetaking course he was running, which became a central nucleating site for the deployment of our prototype for our field study (see §3.3). To continue technical advances that enabled this work as well as augmented collaborative synthesis in his lab (see §5.3), the second author later brought on the fifth author to assist with further technical development.

Data collection and analysis was led by the first author, who led the drafting of the manuscript with elaboration and checks from community-based authors.

### 3.6 Analytic Approach

Analysis was primarily conducted by the first author, using a general constructivist and interpretivist [93] approach. Analysis was interleaved with data collection throughout, including in reflective memos, and, once critical mass of data accumulated, iterative coding and analysis. Analysis was conducted in a theory-informed manner, drawing on sensitizing concepts such as boundary objects [103], sensemaking [88], and infrastructure [64].

## 4 DESIGN OF OUR DISCOURSE GRAPH EXTENSION FOR LOCAL SYNTHESIS WORK

We began our design work building on a previous line of work where we explored how to translate discourse-centric models like SEPIO [18] into a note-taking convention for synthesis. We wrote up this discourse-graph-as-practice translation in [20], shared it with the tools for thought community via a set of video demos and social media posts, and

<b>Design Requirements</b>	<b>Guiding Concepts</b>	<b>Key Components of Discourse Graph Extension</b>
<b>DR1:</b> Enable users to seamlessly create a formal discourse graph alongside their informal notes	Integrated Crowdsourcing (4.1.1), Incremental Formalization (4.1.2)	Node Menu for incrementally authoring discourse nodes alongside informal text (4.2.2); Discourse Grammar Parser for recognizing discourse relations from writing patterns in outlines (4.2.3), Playground for gathering and mapping discourse relations amongst discourse nodes (4.2.3)
<b>DR2:</b> Enable users to leverage their formal discourse graph to improve their thinking and writing	Integrated Crowdsourcing (4.1.1)	Discourse Context for exploring discourse relations between nodes (4.2.4), Synthesis Query for querying discourse nodes and relations (4.2.4)

Fig. 3. Summary of design requirements and design hypotheses developed in Design Arc (Section 4)

tested the process in the first author’s lab with collaborative literature reviews for a few ongoing research projects, as well as a collaborative literature review on COVID with other Roam users. These experiences gave us some confidence that the discourse graph model could be integrated as a *practice* into everyday scholarly work; the remaining question from an infrastructure perspective was **whether it was (socio)technically possible to bridge this discourse-graph-as-practice beyond a single notebook or lab**, to seed new scholarly synthesis infrastructures.

More concretely, we wanted to explore **how we might design a system that made it possible to write and use discourse-graph-as-practice in as natural a manner as possible (e.g., in outlines, with a mix of less formal and more formal argumentation) but seamlessly translate to queryable, shareable discourse graph formalisms in the backend**. Reflecting on our experiences building out these discourse-graph-as-practice working examples, and prior technical and conceptual work, our initial design requirements were:

- (1) **DR1: Enable users to seamlessly create a formal discourse graph alongside their informal notes.**
- (2) **DR2: Enable users to leverage their formal discourse graph to improve their thinking and writing.**

For brevity, we focus our description here on the *outputs* of our design work, in the form of design rationales and initial hypotheses, to frame what we learned from our field deployment and evaluation. Figure 3 shows a summary of these outputs: guided by key design inspirations from previous work (§4.1), we formulated design hypotheses for how to address our key design requirements, embodied in the key components of a discourse-centric software extension for a hypertext notebook (§4.2).

#### 4.1 Guiding Design Inspirations from Previous Work

Our design exploration was particularly informed by the key concepts of integrated crowdsourcing [101] and incremental formalization [96] from HCI research, as well as key insights on synthesis tools (summarized in part in §2.3).

**4.1.1 Integrated Crowdsourcing.** The first design pattern of **integrated crowdsourcing** has been proposed and validated as a means of sustaining work that is beneficial for a community — such as annotating bookmarks on tutorial videos [112] or specifying semantic relationships between ideas on a large-scale innovation platform [101] — that are experienced as tedious work with no immediate individual benefits. The design pattern is to seamlessly integrate the



“crowdsourcing” work (that will benefit a community) into tasks that community members are already intrinsically motivated to do. One example of this design pattern is [101]’ IdeaHound system that integrates semantic judgments of idea relationships in a large-scale innovation platform into the intrinsically motivating activity of visuospatial brainstorming with inspirations from others’ ideas (on computationally augmented digital whiteboards). Another example is Kim et al’s work on “learnersourcing” subgoal labels on tutorial videos by integrating annotation work into the intrinsically motivating activity of creating video bookmarks and notes for oneself [54, 112]. A final example is the Cobi system that sources judgments of research paper relationships for conference planning by integrating these semantic judgments into the intrinsically motivating activity of selecting papers to see at a conference [5, 22].

This design pattern motivated and guided our design work for both **DR1** and **DR2**: our guiding approach was to develop a core user experience that enabled people to write as close to prose as they normally would (and for intrinsic benefits of advancing their own synthesis work), and author a shareable discourse graph as a natural byproduct.

**4.1.2 Incremental Formalization.** The first concept of incremental formalization comes from foundational research on hypertext systems that have explored integrating formal structure — such as schemas — into interactive systems for knowledge management. This line of work has identified how requirements for specifying formal structure — especially if they are the only means of entering data, or imposed early in open-ended tasks — can interfere with the very work they are meant to support [26, 96]. For instance, Conklin and Begeman’s [26] classic work on Group Issue-Based Information Systems (gIBIS), a hypermedia environment for collaborative policy discussions integrated Rittel’s [61] Issue-Based Information Systems (IBIS) framework for structuring collaborative policy discussions, and found that users found the ontology of issues, arguments, and positions, helpful for structuring their discussions, but also wanted ways to integrate this into their work in a more semi-structured way, such as specifying “proto-nodes”, or expanding the ontology to cover more aspects of their thinking, such as ideas and requirements.

Shipman and colleagues [96] proposed the concept of **incremental formalization** as a means of mitigating these risks of formalisms: users enter information in a mostly informal fashion, and then formalize only later in the task when appropriate formalisms become clear and also (more) immediately useful. Notable examples of this pattern include VIKI, a spatial hypertext system that includes heuristic algorithms to find recurring visual/spatial patterns in layout of objects; users can use these to specify schemas if they wish [68]; and the Hyper-Object Substrate (HOS) system, users enter mostly informal text initially, and the system recognizes patterns in the textual information to suggest possible formal attributes or relations for the underlying knowledge base, which the user can then accept/modify/reject as they wish [97]; and the Jourknow system that includes a variety of features that can recognize formal structure (e.g., location, time, meeting information) from (relatively) unstructured notes in pidgin or more lightweight entry format, such as Notation3 [109].

We found this design pattern of incremental formalization to be particularly inspirational for our design work on **DR1**.

## 4.2 The Discourse Graph Software Extension for Hypertext Notebooks

**4.2.1 Hypertext notebook as technical substrate.** Our first design choice was to integrate discourse authoring into the hypertext notebook RoamResearch, for a few reasons.

First, the underlying data model is a graph, which is nicely aligned with the discourse graph model. In Roam, users write outlines, where each bullet (block), and/or page it is contained in, is a node in a graph. Hypertext links (e.g.,

references, embeds, backlinks) between blocks and pages constitute edges in the graph. This data structure was closely inspired by Nelson’s seminal zippered-list data model for hypertext notebooks [78].

Second, the hypertext notebook allows for a mixture of both formal and informal information. Indeed, the base data model in Roam is highly informal and minimally structured: from the user’s point of view, it is just an outline with text. But users can (and do!) add structure through pages, tags, and namespaces, among other things. These are highly beneficial technical properties for implementing the incremental formalization design pattern [96].

Third, as we described in §3.1, the platform has an existing robust and substantial userbase where scholars are already doing work of reading, note-taking and writing. This is an ideal field site for us to test the hypothesis that we can integrate discourse graph authoring into their workflows.

Fourth and finally, the platform is highly extensible, with an API that allows for interactions with and modifications of the underlying data model, as well as the UI. This would allow us to not only implement technical means of authoring discourse nodes and edges, but also prototype UI extensions that could provide intrinsic benefits to the users from the discourse graph being computationally reified. Integrating into existing tools was also motivated by the design pattern of integrated crowdsourcing [101]; we hypothesized that we could develop a software extension to Roam that could integrate the work of discourse authoring into the intrinsically motivating synthesis work that we already knew was happening in these tools.

**4.2.2 Authoring formal discourse nodes alongside informal notes.** Within Roam as a technical substrate, we prototyped a **Node Menu** extension to the UI as a means of authoring discourse nodes in an incrementally formalized manner. The intended feeling was to be similar to highlighting or annotating (rather than forcing structured data entry): we envisioned use of the Node Menu to factor parts of notes — which would be a mix of informal and semi-structured notes — into formal ‘nodes’ of a discourse graph (claims, evidence, etc.).

As an example, Figure 4 (leftmost panel) shows an example of in-progress notes on a research paper, with a general content structure that is similar to a Google Doc of reading notes: a mix of informal and formal observations and structure, including general notes about related ideas, key details about methods, and the core results of the paper. Zooming in on the key results here, a user could mark out the key result of estimated 2x lower susceptibility as a key piece of evidence from the paper that might inform her synthesis for a focal question. The user experience of this would be very similar to annotating: she would simply select the text she wants to “mark as” evidence, press a hotkey to initiate the node authoring action, then select the appropriate node type, either with her mouse or with the associated shortcut (in this case, ‘E’; Fig. 4A).

In doing so, the extension would automatically create a note that is structured with appropriate metadata (and possibly a template) and typed as an “Evidence” (EVD) discourse node, which would appear marked as a formal Evidence node — the prefix “EVD” is automatically prepended to the page title to denote that this is an evidence node, and the @zhuChildrenAreUnlikely2020 citekey is appended to allow for easier return to the source material from the evidence page — inside the previously informal notes (Fig. 4B). The user could stop in her formalization of the nodes at this point, though as her synthesis progresses, and she begins to need more contextual details while comparing and making sense of multiple evidence nodes, she could elaborate the nodes with more details over time, such as by migrating in screenshots of key tables and figures, or methodological details like participants and measures, into the body of the Evidence node page (Fig. 4C).

At the start of our field deployment, our discourse graph extension shipped with the following base set of node types: 1) Questions, 2) Claims, 3) Evidence, and 4) Sources (see Figure 5). Claims and evidence mapped directly to elements

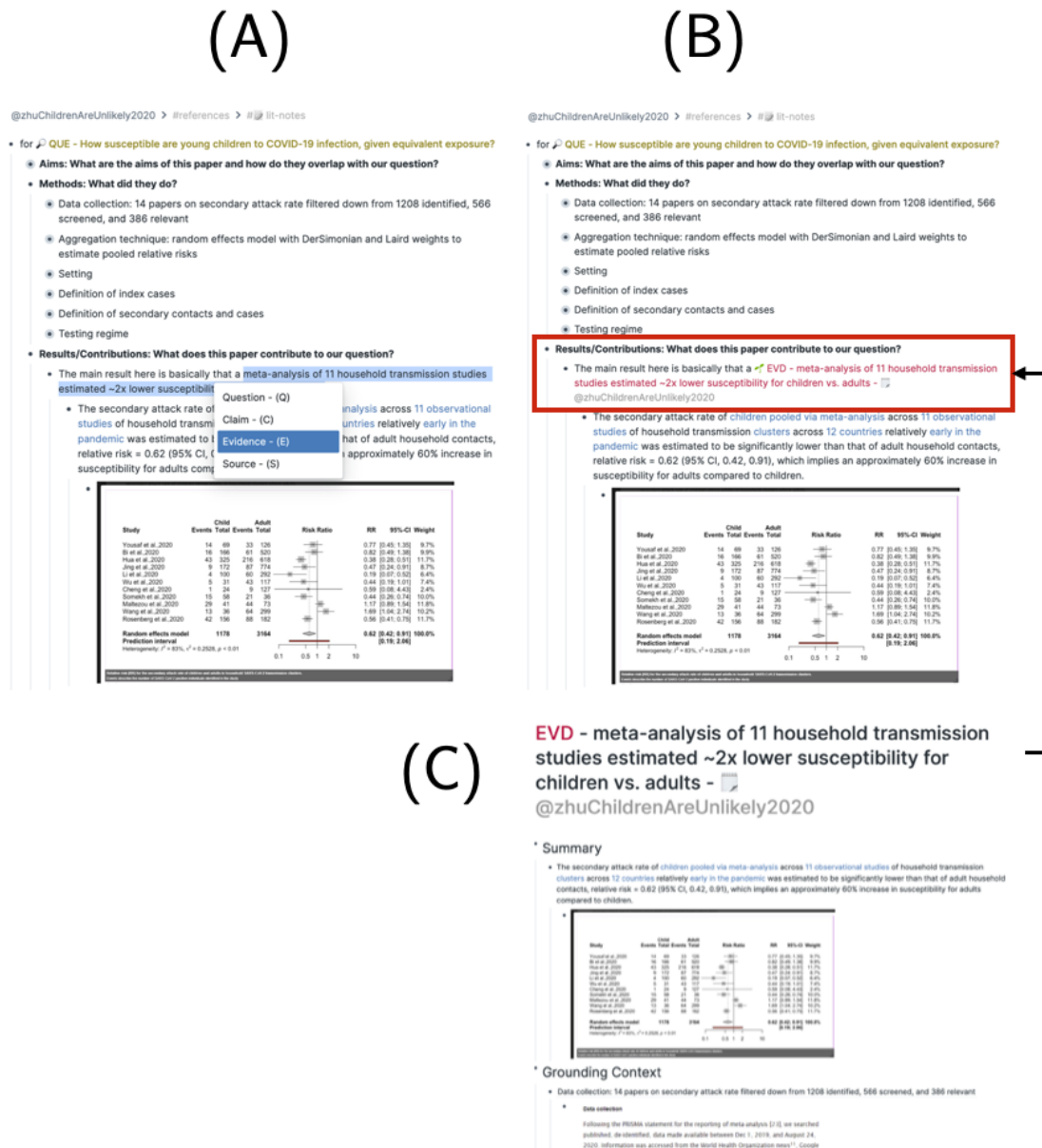


Fig. 4. Our Node Menu component is designed to enable incremental formalization of discourse nodes. Similar to the user experience of annotation, users select the text they want to “mark as” a discourse node, press a hotkey to initiate the node authoring action, then select the appropriate node type, either with the mouse or with the associated shortcut (1). The Discourse Graph extension then automatically creates a page with that title typed as the appropriate discourse node, and replaces the selected text with a reference to that node (2). If needed for further synthesis, users can also elaborate the nodes with more details over time, such as by migrating in screenshots of key tables and figures, into the body of a node page (3).

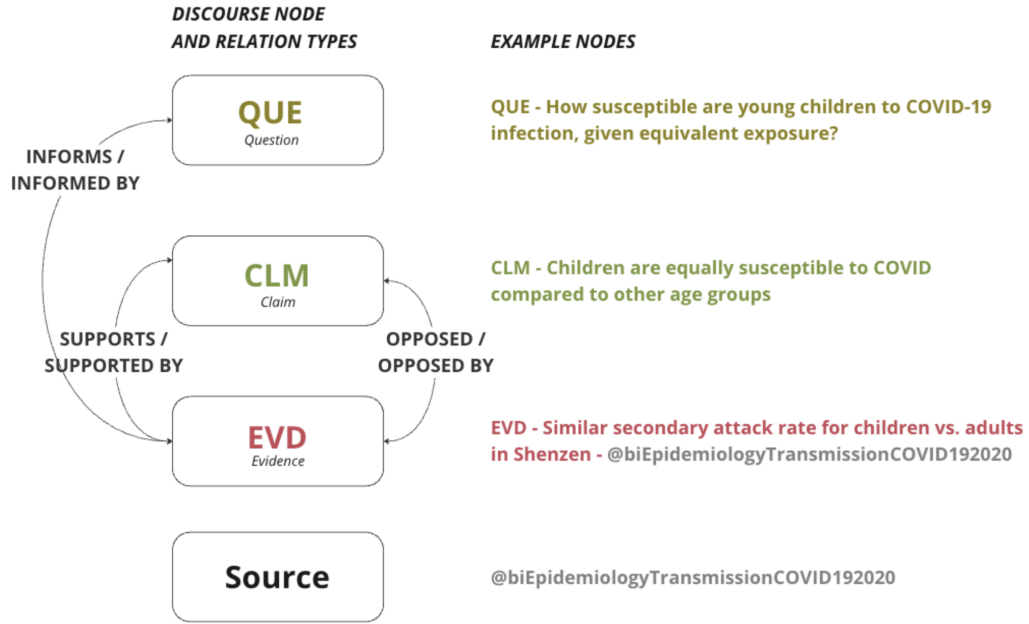


Fig. 5. Base discourse graph grammar that was shipped with the RoamResearch Discourse Graph extension.

of the discourse-oriented models we wanted to integrate with (e.g., hypothesis-evidence from SEPIO [18], claim-data from Micropublications [24]); questions were important from the formative field work for helping to organize and motivate synthesis work; and sources were a bridge to the paper-centric infrastructure of publishing (e.g., all evidence was directly tied to a source, which we hypothesized would enable writing and citing from a discourse graph).

The extension also shipped with a means for users to add new discourse nodes to their grammar (see Figure 25 in Appendix B). This was mostly a debugging feature for the development team, but was also made user-facing, motivated in part by the learnings prior work on incremental formalization that suggested the need for users to expand a formal grammar to structure their thinking while adapting to local conventions and evolving thinking [26], as well as an intuition from infrastructure studies that these seeds of infrastructure should, where possible, design for flexibility to avoid undue path dependence [34]). As we will see later, this design choice enabled user behaviors in the field deployment (see §6.2) that led to significant expansions in our design vision.

**4.2.3 Naturally creating formal discourse relations between nodes by writing and outlining.** To enable seamless authoring of discourse *edges*, we prototyped a **Discourse Grammar Parser** to enable the system to parse formal discourse relation ‘edges’ (e.g., support, oppose, inform) from user-defined conventions of prose writing and outlining. The intent was to make it possible for users to write and outline very similarly to how you would normally write/outline for their own thinking, and author edges between nodes in a formal discourse graph as a natural byproduct of doing this intrinsically meaningful activity (in keeping with the principles of integrated crowdsourcing).

The discourse grammar consists of a set of defined *relation patterns* that map user patterns of writing, which are represented explicitly in a Datomic graph of blocks/pages, and relationships such as referencing and indentation (creating parent/child relationships between blocks), to discourse relations. Figure 6 sketches out the core user flow that

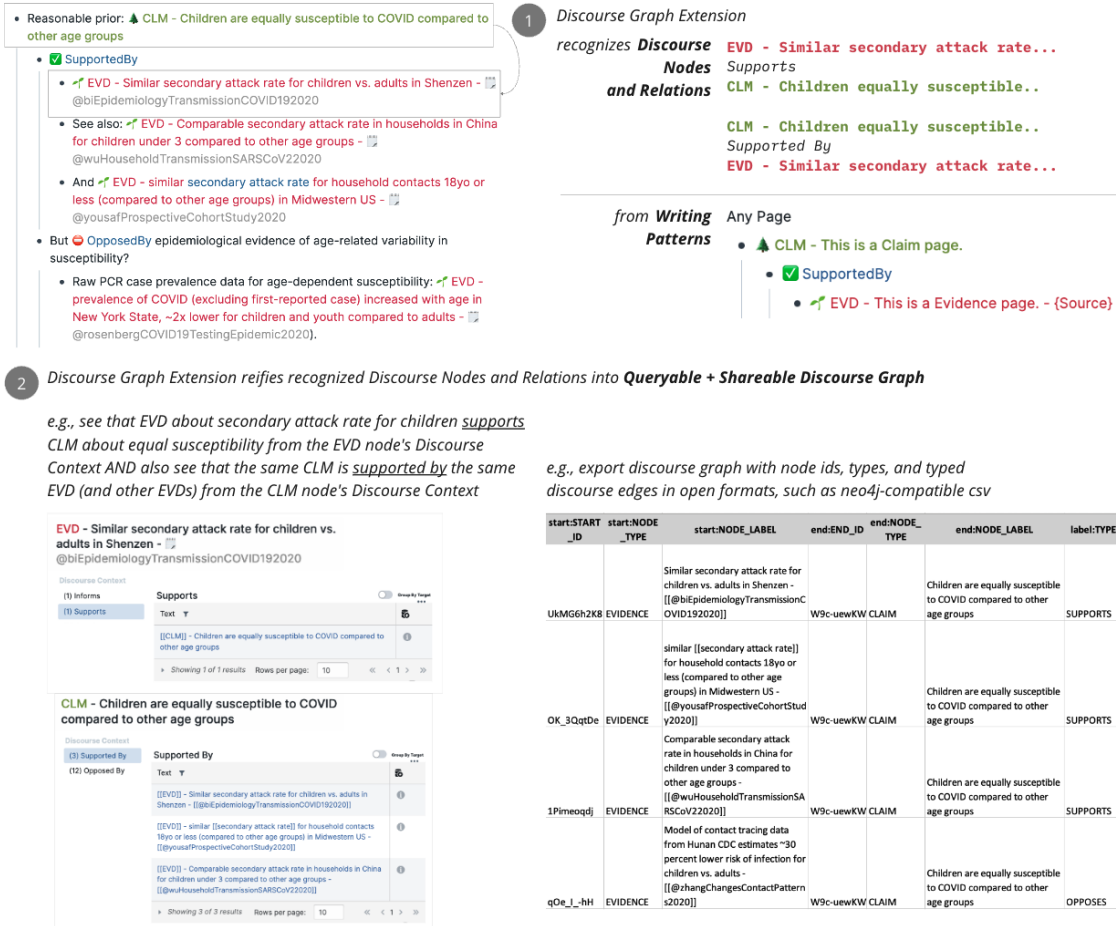


Fig. 6. Our Discourse Grammar Parser parses formal discourse relation ‘edges’ (e.g., support, oppose, inform) from user-defined conventions of prose writing and outlining (1) and reifies the recognized discourse nodes and relations into a queryable and shareable discourse graph (2). This makes it possible for users to write and outline very similarly to how you would normally write/outline for their own thinking, and author edges between nodes in a formal discourse graph as a natural byproduct of doing this intrinsically meaningful activity.

enables this translation. The top left of Figure 6 shows a section of a scholarly outline for synthesis, where a user is writing and outlining her ongoing understanding for a focal question, with a typical mixture of formal and informal notes, and links to resources and references. Here, the user can reference specific results (evidence notes) while making sense of the case for and against a focal claim. Here, each pink EVD link is a link to an evidence node.

In this outline, the first block of text is actually referencing a **claim** that children are equally susceptible to COVID compared to other age groups. This is done through a linking feature in Roam that is not so different from hyperlinking in Web-based communications, such as email. A child bullet (or block, in Roam parlance) references an **evidence** node from Bi et al (highlighted in pink); in this way, the inline citation to Bi et al is actually to a specific result from that paper in an evidence node, rather than a reference to the whole paper. We also see references to pink evidence nodes

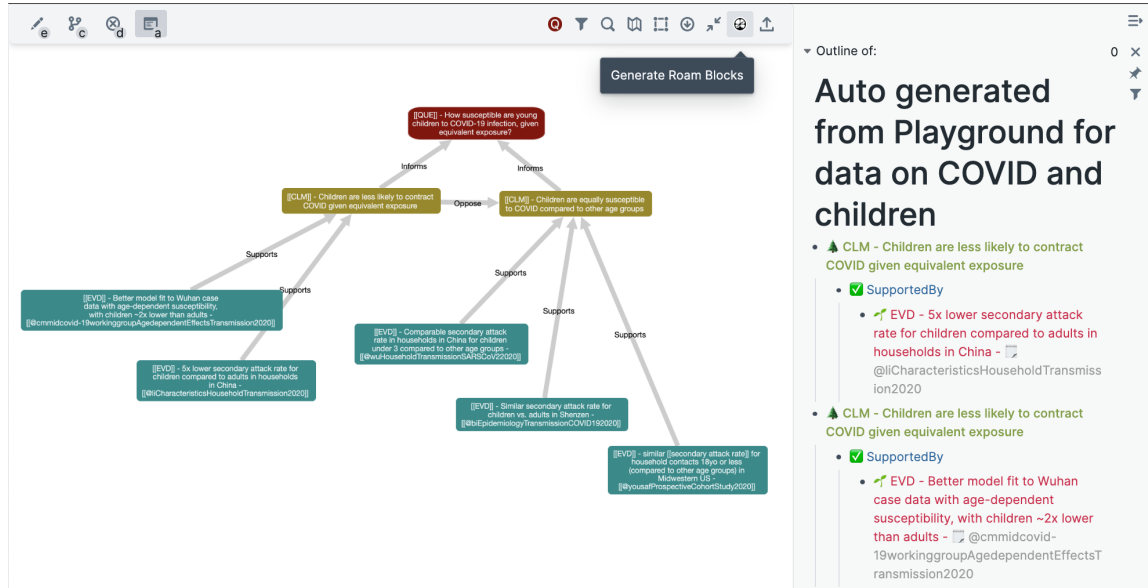


Fig. 7. Our Playground interface provides an infinite canvas component where users could insert or create discourse nodes, visually cluster them, and draw edges between nodes; they could then make the nodes and edges "real" by triggering the parser to generate the associated Roam blocks in the formats that specify the discourse nodes and edges according to the relevant relation patterns defined in the grammar.

in the text. To record and reason through the evidence that supports the claim, the user could note in writing in the second block that the claim is "Supported By" some evidence, by referencing a tag with the title "Supported By". If so, the discourse graph grammar could then recognize the formal discourse relations amongst these nodes (see Figure 6, Step 1), and reify them into a queryable and shareable formal graph, from which the user can reap immediate benefits, e.g., by exploring and querying their discourse nodes and relations (described in §4.2.4; see Figure 6, Step 2). Indeed, this translation of implicit, textual specifications of discourse relations in writing would be in effect for all the discourse nodes referenced in the synthesis outline. Figure 6 (bottom right) shows a portion of the exported subgraph of discourse nodes and relations that could be parsed from the writing in the outline, in a Neo4j-compatible csv export from the Roam notebook.

Separately, inspired by previous work on visuospatial sensemaking and visual hypertext systems (e.g., IBIS [26]), we also prototyped a Playground infinite canvas component where users could insert or create discourse nodes, visually cluster them, and draw edges between nodes; they could then make the nodes and edges "real" by triggering the parser to generate the associated Roam blocks in the formats that specify the discourse nodes and edges according to the relevant relation patterns defined in the grammar.

At the start of our field deployment, our discourse graph extension shipped with the following base set of discourse relation types, focusing on relationships between evidence and claims/questions:

- Evidence **INFORMS** Question
- Evidence **SUPPORTS** Question
- Evidence **OPPOSES** Claim

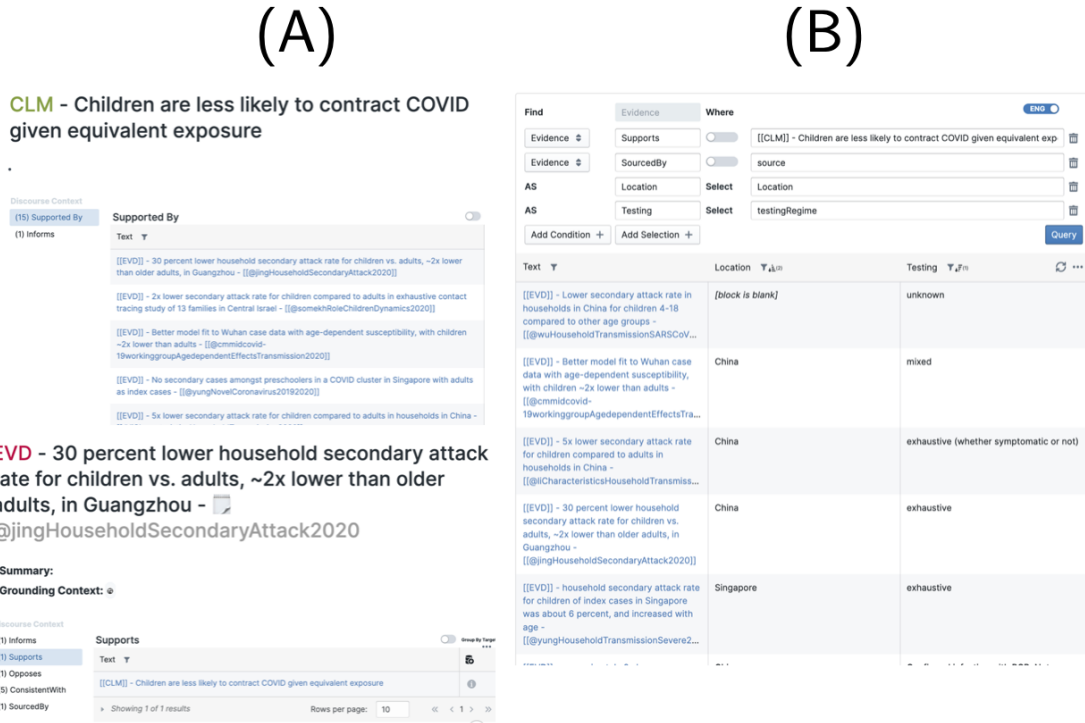


Fig. 8. Our Discourse Context component was designed to provide associative trail navigation between notes based on discourse relations like inform/support/oppose (A). We also prototyped a Synthesis Query component to enable users to perform specialized querying based on discourse graph elements and relations (e.g., find all evidence that informs my question, but opposes some claim; find all evidence that informs a question and display key methodological attributes for each piece of evidence that I have specified (B). These queries could be used as collections, or as a means to gather discourse nodes to insert into a Playground or outline for synthesis.

As with the node types, the extension also included a means for users to edit the in-built relation patterns, as well as add their own to match their own styles of writing and outlining (see Figure 25 in Appendix B).

**4.2.4 Enabling users to derive immediate benefits from their queryable and shareable formal discourse graph.** The final components of our design were designed to enable users to reap immediate benefits from their local discourse graph.

We hypothesized that having a formal discourse graph in a hypertext notebook could help with synthesis in many ways, such as querying and sorting between claims based on evidentiary support/opposition. To explore this, we prototyped a **Discourse Context** component that provided additional associative trail navigation between notes based on discourse relations like inform/support/oppose (see Figure 8A). This discourse context was designed as a slightly more formalized/focused version of the native Roam backlinks feature that enabled users to traverse bi-directional links between notes. Importantly, because the discourse relations are computationally 'real', discourse relations can be traversed from source to target, or vice versa, without requiring the user to tediously specify both directions of a discourse relation (e.g., EVD Supports CLM, CLM SupportedBy CLM): as Figure 8A shows, once the parser has reified a

discourse relation between nodes, such as a "Supports" relation between an EVD and a CLM, the user can see the EVDs a CLM is SupportedBy (from the CLM) page, and also navigate from any given EVD back to the CLM that it Supports.

We also prototyped a **Synthesis Query** component to enable users to perform specialized querying based on discourse graph elements and relations (e.g., find all evidence that informs my question, but opposes some claim; find all evidence that informs a question and display key methodological attributes for each piece of evidence that I have specified; see Figure 8B for an example). These queries could be used as collections, or as a means to gather discourse nodes to insert into a Playground or outline for synthesis.

## 5 EVALUATION I: INTRINSIC BENEFITS OF DISCOURSE GRAPHS FOR LOCAL SCIENTIFIC PRACTICES

Our deployment yielded a rich picture of a possible future of a discourse-centric infrastructure being grown from transformed local scientific practices: a thriving ecosystem of personal and lab hypertext notebooks that researchers use to author local discourse graphs because they improve synthesis (§5.1), enhance primary research and research training (§5.2), and augment collaborative research (§5.3).

### 5.1 Augmenting Local Synthesis Work with Discourse Graphs

*5.1.1 Opening Vignette.* LZ is a PhD student in medical image processing, based in Shanghai, China. He came to know about the extension and approach via the initial whitepaper we published on the synthesis workflow [20]. He learned how to use the extension specifically from one of the Roam notetaking courses in which we deployed the extension, as well as tutorials and examples from Youtube videos and other resources from the community of users.

He had been taking notes in Roam for awhile, but without much explicit structure. In one of his main writing projects, he was struggling to conceptualize the contribution of his work against the existing literature, and the extension enabled him to write out and trace the "logic line" that articulated the contribution of his work, enabling him to write a compelling introduction to his paper. In an interview, he shared<sup>6</sup>:

*In my former studies...I read from about dozens of papers Then...I find that this logic line...I find that, and through another one and through another one, I find this is the... Oh, my God! If I not[sic] write it down, all these lines, my my brain will not understand this this story, like there's a line...this is amazing!*

His notes on the questions, claims, and evidence were written in outlines and rough scratch notes throughout his notebook, but he was able to use the extension to gather the major discourse elements together to visually map out the "logic line" for his paper visually on the Playground (see Fig. 9A)<sup>7</sup>. We can see here how the project's core argument and motivation about the scientific problem of informative frame selection of laryngeal cancer is mapped out in a series of core questions and claims, from the severity of laryngeal cancer, to its carcinogenesis, to motivating early diagnosis as an effective approach to addressing this disease, and the specific technique of narrow band imaging that is studied in the project.

He shared in an interview in September 2023 — following up approximately 1 year after he responded to our usage survey — that he was still actively using the extension for his work, and has used the extension to assist with mapping out the argument for an additional publication. More generally, the "bread and butter" of his work with the extension (aside from intensive writing for a focused publication) is heavily question-centric: he writes regularly in his "daily

<sup>6</sup>Zoom call 2023-09-29

<sup>7</sup>Graph snapshots shared by LZ 2023-10-20



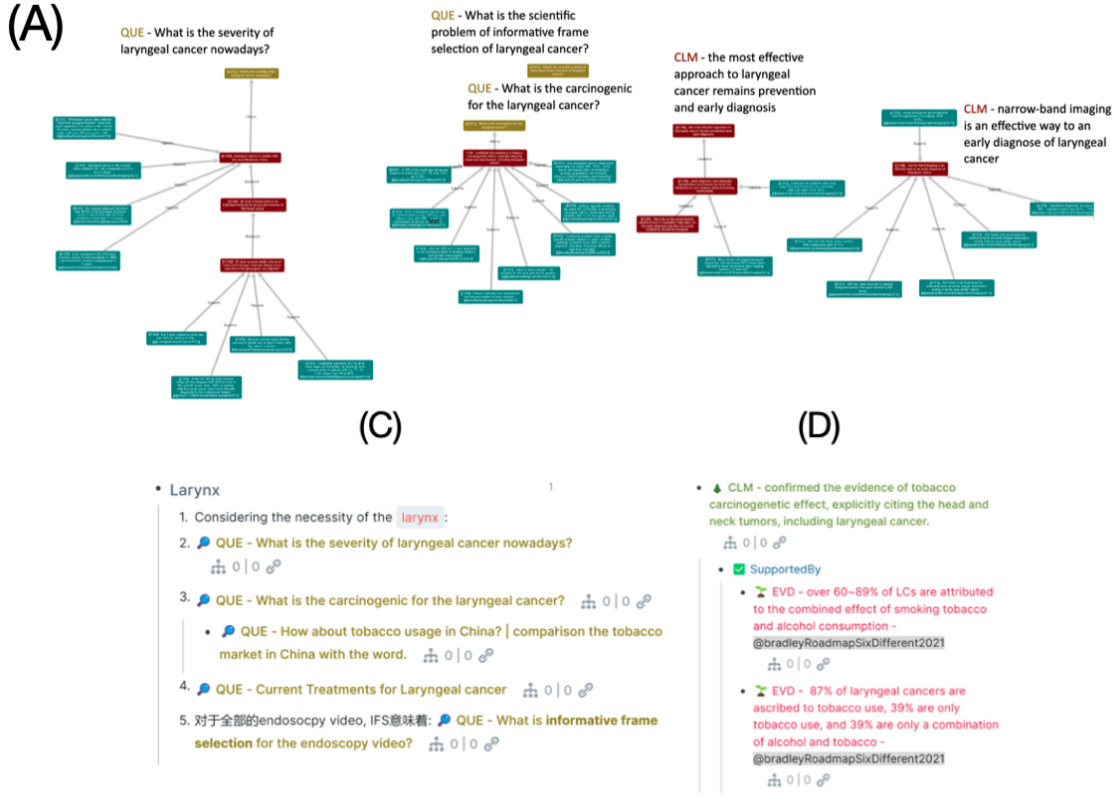


Fig. 9. Snapshots of LZ's usage of the DiscourseGraph extension. (A) shows how LZ visually maps questions (gold boxes), claims (maroon boxes) and evidence (teal boxes) in the Discourse Graph Playground to make sense of the "logic line" of a research project's contribution to the literature. The snapshot is annotated with the text of key questions and claims in the argument. (C) and (D) show LZ's question-centric usage of the DiscourseGraph extension: he tracks questions for topics of interest, and elaborates on them with relevant claims and evidence. Here, (C) shows a series of questions around the importance, causes, and diagnostic approaches and issues for laryngeal cancer, and (D) shows elaboration of the question of the carcinogenesis of laryngeal cancer with relevant claims and evidence.

notes", noting down new questions that motivate him, or elaborating on existing questions with new sources, claims or evidence (see Fig. 9 C and D)<sup>8</sup>.

**5.1.2 Using Discourse Graphs to Plan and Produce Research Artifacts.** LZ's usage of discourse graphs to plan and produce research artifacts was echoed by a number of other participants. For example, as of Fall of 2023, AP had used the extension to research a chapter to be published in an academic collection of essays, and also to prepare a compilation of positions, extracted from statements made by governments at the United Nations in the last several years, to serve as a basic guiding paper for negotiations on a potential international treaty regulating international cooperation in disaster relief assistance and risk reduction. He shared that using the extension to see "some of the patterns of positions was particularly useful"<sup>9</sup>. Another striking example was MB, who used the extension to deepen and speed up the drafting of

<sup>8</sup>Field notes and Zoom call 2023-09-29

<sup>9</sup>Discord message 2023-09-25

his dissertation, including being able to write a draft of the first chapter of his dissertation (the introduction) in less than 2 days. He described the mechanics of it like this on a user call in terms of enabling him to map out and think through the structure of his thinking sufficiently that putting words on the page for a draft became much easier<sup>10</sup>:

*"So when you go to writing I'm like, I'm not going to pick up that article again. I'm going to write from here, so if it's not here I'm in trouble right. So my first paragraph might be, there's a vast amount of theories that you know could supply the literature: Boyd discusses bargaining power, Bertrand discusses, you know, parental attribute theory, Gomez discusses whatever Gomez discusses right um. And boom paragraph done just because I have, you know, put this together."*

Similarly, GX shared in a usage survey that the extension enabled him to *"easily sort out the framework and context of the article I want to write"*, especially because he was branching out from his prior background in public relations to criminology for his PhD work<sup>11</sup>.

**5.1.3 Using discourse graphs to address specific process issues in synthesis work.** From our user calls and discussions and surveys, we got more of a picture of how the extension enabled users to **improve their synthesis work by addressing specific weaknesses or desire paths in their synthesis process**. One common dimension of this was mitigating issues from relying solely on memory for details and context. MB, elaborating more on his usage of the extension, described how he was able to *"pull things up in seconds"*, in contrast to his peers in his comprehensive exam course, who often used programs like Word, and struggled to pull up details over the course of minutes<sup>12</sup>. Additionally, in a call describing how the usage of the extension was impacting his work, MA said that he appreciated how the clearer documentation of takeaways from previous papers in terms of claims and evidence enabled him to rely less on only his memory from reading papers, and instead be far more rigorous and justified (and less wasteful, not running simulations on questions that already have good answers) when thinking and planning out parameters for computational models<sup>13</sup>:

*"[M]ore or less previously I was only sort of remembering as much as I could from the literature, from like an initial read through the literature: what's, what's, what's been done and what hasn't. And and I was running simulations that I thought would be interesting, but it wasn't rigorously justified that.. let's say I'm varying parameter X, maybe parameter X is already known from the literature, but I, I just didn't read the literature carefully enough."*

Similarly, in our usage survey, PO, a clinical researcher, said (in response to the question "Why did you start using the extension? What value is it providing to you now?"<sup>14</sup>:

*"I read a lot of medical papers. A lot of times I don't remember who wrote this, how I reference that ... where did I read that ... I thought it would be an excellent tool to organize knowledge."*

And QQ — a physics PhD student — noted, for the same usage survey question about value<sup>15</sup>:

*"I need to connect the claim with both research question and literature. I think it is helpful for me to check which literature it comes from."*

Another prominent theme was the usage of the extension to **increase the signal to noise ratio in their personal knowledge bases**. For instance, AP said, in response to the usage survey question about value<sup>16</sup>:

*"it has helped me apply a metadata "language" to the links in my graph such that I can see patterns and relationships between different nodes (containing information), that would simply not be discernible from the "noise" of thousands of links."*

<sup>10</sup>Open office hours call 2022-04-07

<sup>11</sup>Usage survey GX

<sup>12</sup>Field note 2022-03-11

<sup>13</sup>Zoom call 2021-11-05

<sup>14</sup>Usage survey PO

<sup>15</sup>Usage survey QQ

<sup>16</sup>Usage survey AP

Similarly, in response to the same question, CV, an independent sociologist, said <sup>17</sup>:

*"This extension responded to my need to overcome the fragmentation of pieces of information that I had in Roam (pages named after a single concept, different words to refer to similar concepts, etc.). To solve this problem, Discourse Graph has been a wonderful tool, because it has allowed me to organize the information on a more aggregate level than the one allowed by the use of hashtags/pages."*

And BO, a qualitative social researcher, described how the Discourse Context allowed for higher-signal ways to explore relations to other nodes and ideas compared to the in-built "backlinks"<sup>18</sup>:

*"The 'Discourse Context' section allows me to focus on important relations between my notes. In the Linked References section there is much more 'noise'."*

This comment about the Discourse Context providing a higher signal to noise ratio than the native backlinks feature is consistent with observations of usage, as well as data from the usage survey. In response to a question about the extent to which various DiscourseGraph extension features were currently in use (with response options ranging from "Didn't know about this", "Never used", "Use on occasion", and "Core part of my workflow"), 70% of respondents who self-identified as current active users of the extension reported that the Discourse Context feature was a core part of their workflow, and 17% reporting that it was used at least on occasion. This was by far the most used extension feature (with the closest second being the Node Menu: 57% reporting that it was core to their workflow, and 26% reporting that it was used on occasion).

One key design iteration that happened during the field deployment sheds further light on the potential of integrated discourse graphing to address specific weaknesses or desire paths in users' synthesis processes: we prototyped a **Discourse Overlay** feature that enabled users to see an in-line preview of the discourse context of a discourse node — in the form of a count of the number of discourse relations for a discourse node, as well as the number of references to that node, appended to the right of each reference to that node in an outline — as well as explore the discourse context for the node — by opening a "popup" overlay that displayed the **Discourse Context** for that node — without leaving the context of a larger outline or argument they were being used in. The "ratio" of discourse relations to references was inspired by the citation "badge" from Scite.AI that shows separate counts of "higher signal" Supporting or Opposing citations to a paper, relative to the number of "mentioning" citations, to enable more nuanced reasoning over the impact of research papers than just raw citation counts.

One core goal — in response to user feedback — was simply to increase the accessibility of the Discourse Context of each node, enabling more efficient retrieval and traversal/exploration of these "higher signal" relations and ideas. One comment from a user — who was collaborating with the first author on a project — illustrated this dimension of value, describing the Discourse Context "popup" as extremely helpful for quickly exploring the discourse context around a set of claims and questions in an outline (see Figure 10A)<sup>19</sup>. Another core goal — also in response to user feedback about wanting to be better able to leverage their discourse nodes in their synthesis — was to enable more "talkback" [94, 116] from the discourse nodes to guide and support development of ideas: for example, discourse nodes that are heavily used or referenced in an argument but with very few discourse relations might represent ideas that need further development; nodes that have many discourse relations might indicate well-developed ideas that are more ready for use in synthesis. The first author used this in his own writing and outlining for this exact purpose, to keep track of the "strength" of different pieces of his argument while outlining with discourse nodes (see Figure 10B). This value proposition was echoed by AP, who used the overlay extensively during synthesis, and shared this comment with us<sup>20</sup>:

<sup>17</sup>Usage survey CV

<sup>18</sup>Usage survey BO

<sup>19</sup>Discord message 2022-01-04

<sup>20</sup>Discord message 2022-01-14

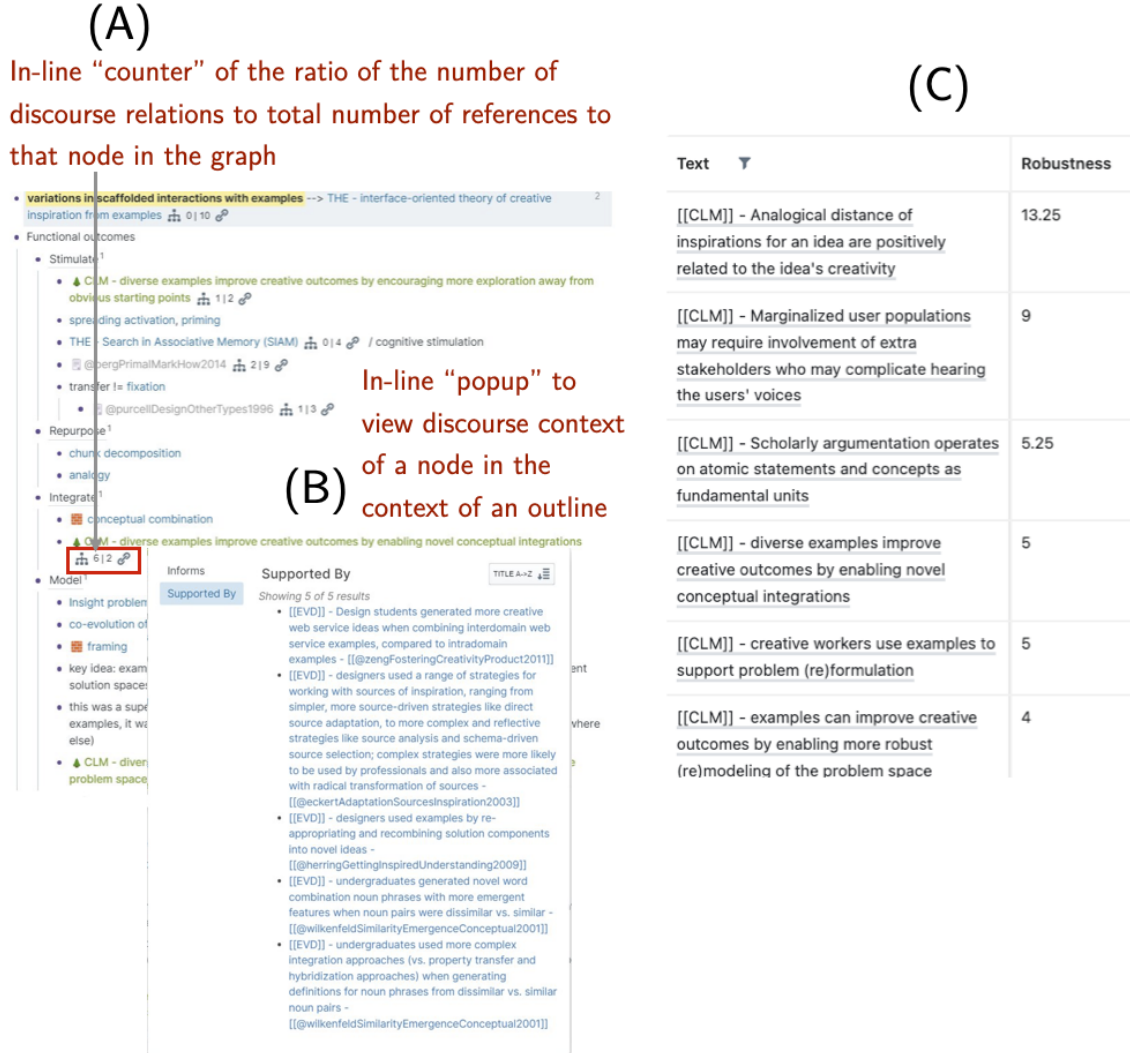


Fig. 10. Snapshots of usage of the experimental Discourse Overlay and Discourse Attributes features for enhancing users' ability to leverage discourse nodes and relations in their synthesis work. (A) Using the inline ratio of discourse relations to number of references for a discourse node to keep track of the "strength" of different pieces of an argument while outlining with discourse nodes. Specific discourse relations and nodes are easily accessible via the overlay popup of the Discourse Context for each node (B). An experimental "discourse attributes" feature allows users to define named mathematical functions over the discourse relations of a node (e.g., weighted sum over number of supporting claims, evidence, and sources) that could be displayed in the overlay and used in queries, such as showing the most "robust" claims in a graph (C).

"[I]nitially the ratio did not register with me. However, as I started using the overlay more and more I started finding myself opening up claims etc that had interesting reference-to-link ratios. It is the "pull" or "weight" you refer to. It is really powerful and is at the core of the d/graph. You have set up a complex system, and it deserves a more attenuated or granular representation to really convey the power of the underlying relationship structure that is being developed"

We also prototyped an enhancement of this overlay to enable specification of custom **Discourse Attributes** to display in the overlay and use in queries of the discourse nodes (see Figure 10C). These discourse attributes could be a mathematical function of a node's discourse relations: for instance, one could define the "robustness" of a claim as a weighted sum of the number of supporting claims, evidence, and sources, perhaps also including a "penalty" for opposing claims/evidence.

Unfortunately, the overlay was very performance-intensive, and we were unable to find workable solutions for optimizing it, given the limitations of the querying mechanisms we had technical access to for the extension, which led to very noticeable and disruptive slowdowns in the user interface. These slowdowns led to the feature becoming mostly left unused due to performance issues. Even here, though, user reports of these performance issues underscored the value of the feature. AP's comment from the usage survey (in response to the question "What are your biggest pain points or wishlist items at the moment with the extension?") is particularly illustrative<sup>21</sup>:

*"I have stopped using the d/graph overlay because it slows my graph down significantly (an issue that has been there from the beginning). It would be great to have a solution for that because I have found that accessing the traditional context menu add extra clicks, i.e friction, which slows things down. After setting it up, there are two components of working: adding d/graph nodes and then afterwards processing the information from the d/graph (i.e. either looking at the relationships in the context menu or doing queries). The former is much easier/more efficient than the latter. While search/querying is powerful, it is sometimes overkill. If the overlay worked on large graphs/questions pages it would be a really good intermediate solution."*

## 5.2 Improving and Transforming Primary Research Work with Discourse Graphs

**5.2.1 Opening Vignette.** MA is a cell biologist who came to learn about the extension and discourse graphing approach while preparing to start up a new lab as a tenure-track faculty member at an R1 university in the United States. He uses computational modeling and fluorescence microscopy of stem cells to study cell biology and biophysics questions, like the mechanisms of cellular membrane bending by self-organizing cytoskeletal protein assemblies. The goal is to make experimentally verifiable, predictive computational models of self-organizing processes inside cells.

One motive for usage was to better bridge his computational theoretical work with experimental literature. In his experience, there was often very little formal connections between the two, such that modeling was not systematically backed by experimentally derived knowledge, limiting productive intercommunication between models and experiments. A concrete example of this: in models of the role of type-I myosin in endocytosis, MA needed to choose and study plausible value for such parameters as the binding radius, binding rate, and length of myosin-I. He used the extension to systematically map and collate experimental observations of this angle to inform his modeling work.

Very quickly, however, he began to adapt the extension to integrate into his day-to-day experimental work in simulations and at the bench. He realized that he could adapt the ontological distinction between claims and evidence to his day-to-day work of forming and testing *hypotheses* against experimental *results*. This enabled him to retain rich discourse context for his inquiry, tracing his original experimental results directly to the context of past claims and evidence that bear on the same overarching question. He was able to extend the discourse grammar with new nodes (HYP (hypothesis) and RES (result) nodes to formalize this extension) using the grammar editor. Figure 11 shows a snapshot of how he used the DiscourseGraph extension to structure inquiry from literature and primary experimental work for a question about the stoichiometry of Hip1R binding to actin filaments. In the playground (Fig. 11C) he is tracking an overall question about the stoichiometry of Hip1R binding to actin filaments. He is questioning a specific claim

<sup>21</sup>Usage survey AP

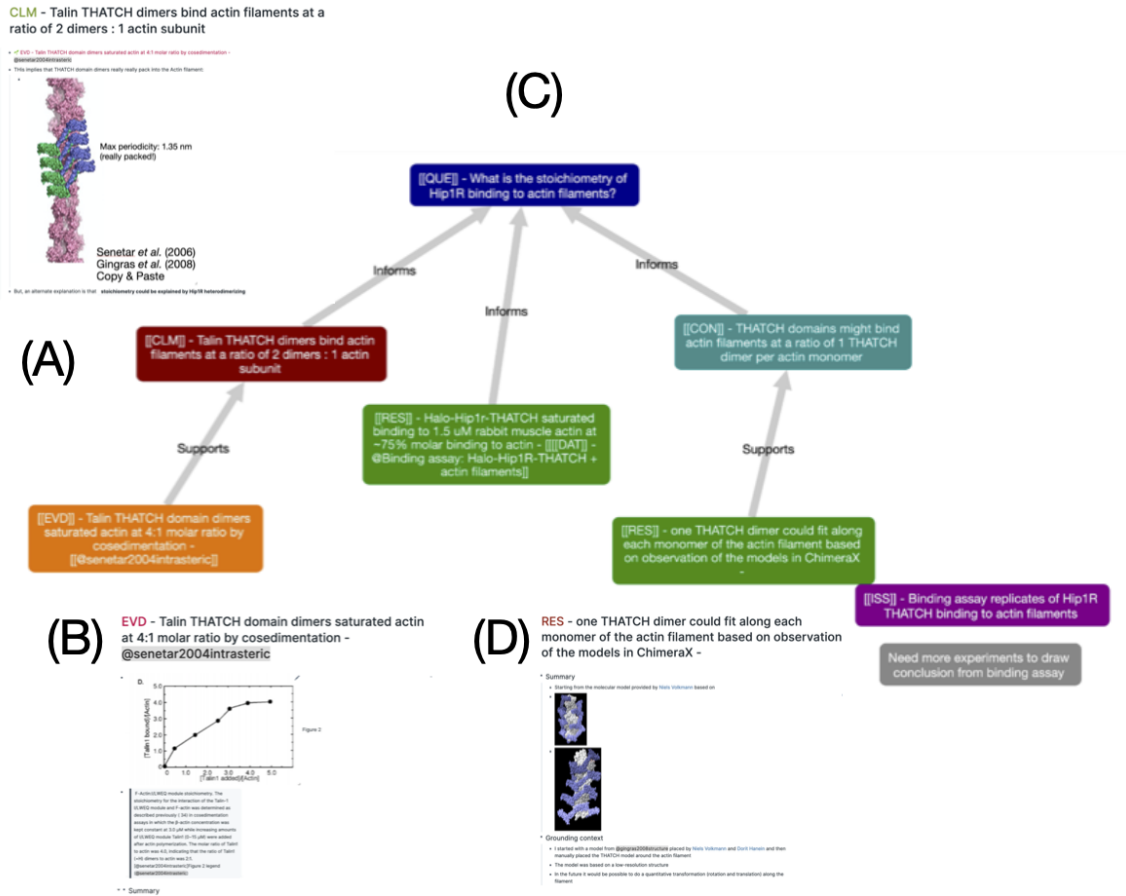
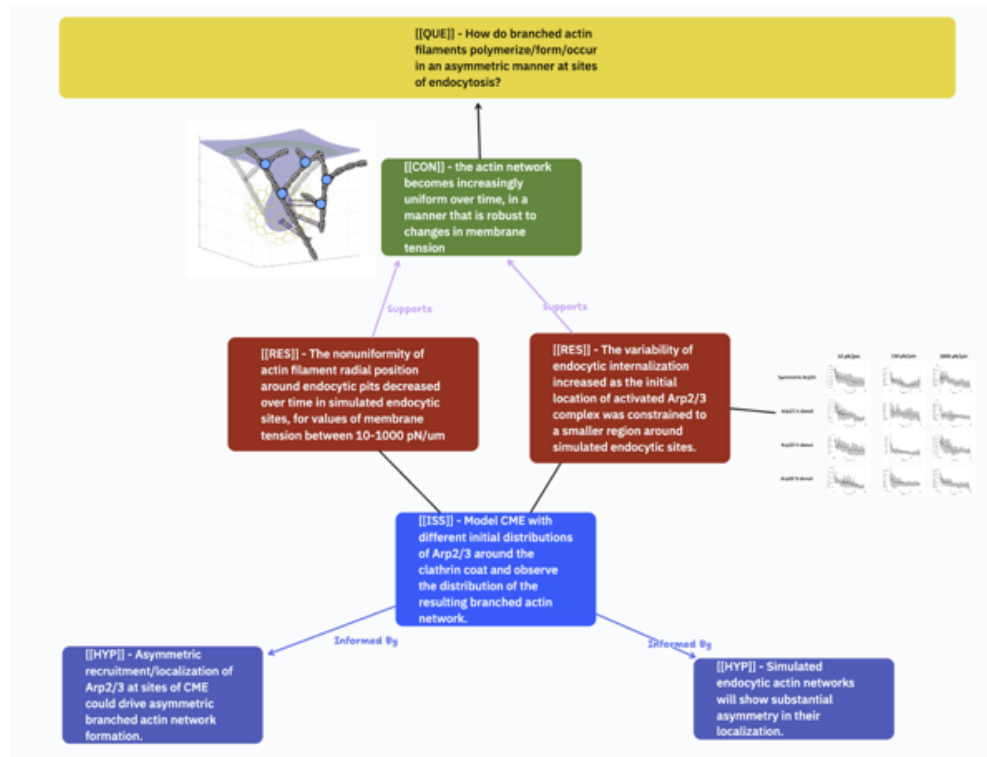


Fig. 11. Snapshots of MA's usage of the DiscourseGraph extension for structuring synthesis and primary research for a single project. (A) and (B) show a Claim and Evidence node from the literature that informs a driving question about the stoichiometry of Hip1R binding to actin filaments. This prior discourse is shown in context of a Playground snapshot (C) alongside Results and Conclusions for the same driving question. Note how the playground also includes an "Issue" node that describes an experiment that needs to be run to continue the inquiry, and a freeform note about what other experiments are needed. (D) shows the contents of a Result node about how one THATCH dimer could fit along each monomer of the actin filament based on observation of the models in ChimeraX.

about the ratio being 2 dimers to 1 actin subunit, based on a specific result from a prior study (indicated by the citekey @senetar2004intraasteric) that "Talin THATCH domain dimers saturated actin at 4:1 molar ratio by cosedimentation". The EVD node itself (11A) contains a screenshot of the key figure that depicts the result, as well as a snippet of text from the paper that describes the result. MA has also elaborated on the specific methods context of the result ("muscle beta actin was used"), as well as some freeform notes about possible interpretations. Note also the new HYP, RES, CON, and ISS nodes that are added to the grammar to enable this integration of synthesis with primary research. From the structure, it is clear that HYP and CON are the primary research analogs to CLM nodes, while RES is the analog of EVD nodes.

(A)



(B)

I am a	I am looking to	Start here
<b>Researcher</b>	get up to speed on the answer to a particular research question	<a href="#">QUE Questions with a Summary</a>
Researcher	find an open research question	<a href="#">Open QUE Questions</a>
Researcher	find testable hypotheses	<a href="#">HYP by topic</a>
<b>Student</b>	contribute with limited domain knowledge	<a href="#">EVD Evidence that needs more context</a>
Student	choose a simulation or analysis to run	<a href="#">ISS by topic</a>
<b>Generalist</b>	summarize the answer to a research question based on collected results	<a href="#">QUE Questions that need a Summary</a>
<b>Established researcher</b>	see recent research results	<a href="#">RES by topic</a>
Established researcher	share caveats about a given experimental result	<a href="#">EVD Evidence that needs more context</a>

Fig. 12. Snapshots of the usage of the DiscourseGraph extension for structuring work in MA's cell biology lab. (A) shows how the extension enables a succinct documentation of the hypotheses, experiments, results, and conclusions of a single quarter rotation by a rotating PhD student, enabling modular attribution and credit tracking. (B) shows the construction of an onboarding dashboard that uses discourse nodes as entry points for contributors of varying levels of expertise, established researchers (who can find open research questions or testable hypotheses or specific experiments they can contribute to), to generalists (who can summarize what is learnt for a given QUE (question)), to students (who can elaborate on evidence nodes or run a scope simulation or analysis).

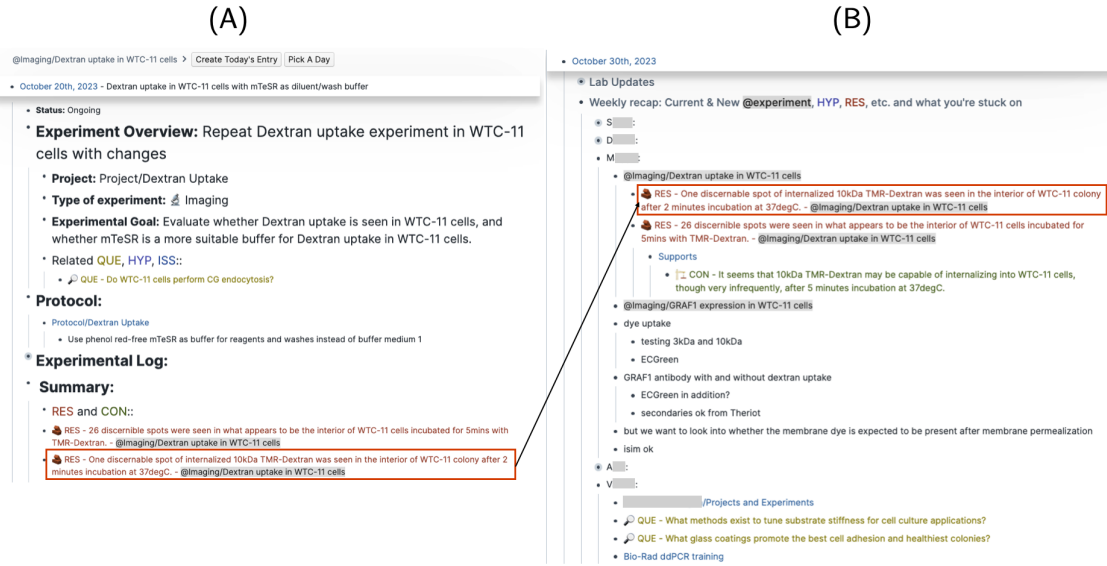


Fig. 13. Snapshots of MA's lab's usage of the extension to structure primary research: (A) shows a recently completed section of an experiment documented in MA's lab graph by a lab manager, documenting the questions that motivate the experiment, and the tangible outputs of the experiment summarized as result nodes. (B) shows how the result nodes are then shared and discussed in lab meeting notes.

MA has since deeply integrated discourse graphs into the practice of his new lab, which he started up in the Summer of 2022. At the time of this writing, his lab includes 1 research scientist and lab manager, 2 PhD students, and 3 undergraduate students. MA is using the extension and the shared Roam graph to implement his vision for a more deliberate, nonhierarchical approach to lab science, that emphasizes empowerment and ownership of work by all participants, and a pace and openness of work that opposes the prevailing publish-or-perish hypercompetitive culture of academic science. His overall theory of change is that new collaboration tools can facilitate cultural and organizational change, and that using discourse graphs to break research into modular parts that can be shared and cited could facilitate getting up to speed on a new field of research, identifying holes in research to motivate new experiments, create new theories of understanding that synthesize interdisciplinary data, and regular exchange of discretized results between researchers. One concrete use of the extension for this goal was to more concretely structure and document contributions of students who rotate in for a single quarter (see Figure 12A). MA also hired a research assistant "cybrarian" to specifically focus on the work of structuring the lab graph and developing and enforcing (software) conventions, including the onboarding dashboard shown in Figure 12B that describes entry points into contributing to the lab's science for researchers of varying expertise and experience levels.

Figure 13 shows how the lab now uses individual experiment pages — analogous to individual papers, though yet to be published — , where questions are motivating the experiment, and the tangible outputs of the experiment are summarized as result nodes (Fig. 13A)<sup>22</sup>, that can then be shared and discussed in lab meetings (Fig. 13B)<sup>23</sup>.

<sup>22</sup>Page 2\_Yv78JF in MA's lab graph

<sup>23</sup>Page iGKQlvIO- in MA's lab graph



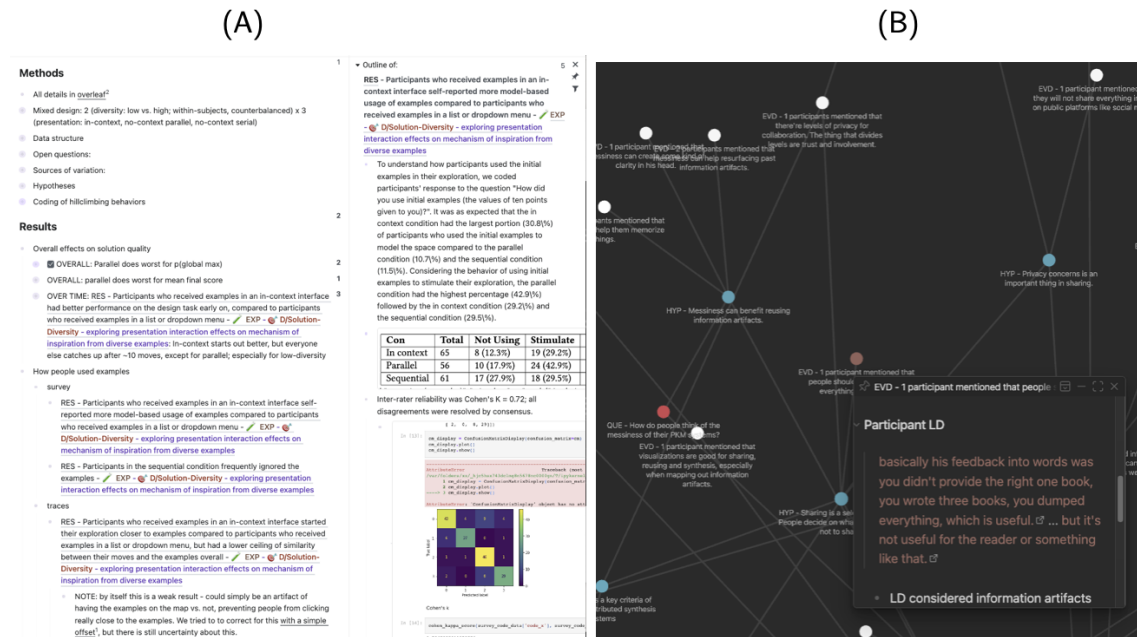


Fig. 14. Snapshots of the first author's lab's usage of the extension to structure primary research: first for tracking the soldiv experiment results (A), and then for structuring ongoing qualitative analysis in terms of Hypotheses and Evidence (with links to qualitative data points).

**5.2.2 Using Discourse Graphs to Structure Ongoing Primary Research.** MA's vignette illustrates a surprising observation from the deployment: researchers could integrate the DiscourseGraph extension to not only structure synthesis of prior literature, but also structure ongoing primary research. This was not surprising from a logical information model perspective: the information models that are translated in the Discourse Graph workflow, such as micropublications [24] and nanopublications [59], were after all proposed as an alternative model for reporting on primary research (instead of traditional narrative publications); as such, the elements of the base grammar map readily to ongoing primary research (e.g., a claim could be a "hypothesis" in development, and evidence could be a "result" from an experiment that was just completed). The surprise came from our initial sense that the workflows and practices for writing up primary research were too entrenched to change.

The first author has since adopted this ontology in his own work, starting with tracking documentation of results from an experiment and then thinking through how to sequence those results into a coherent larger contribution (Fig. 14A). More recently, the first author and one of his students have been using the discourse graph workflow to structure analysis of a qualitative study (Fig. 14B)<sup>24</sup>; it should also be noted that the analyses for this paper itself were also structured in a discourse graph workflow, albeit in a different hypertext tool, Obsidian, and independently initiated by the student (more on this in §6.3). A related desire to use the extension to support primary qualitative research was also independently expressed by GM, an independent researcher, on the Roam Discord<sup>25</sup>:

<sup>24</sup>SIF analysis graph

<sup>25</sup>Discord message 2021-11-20

*Are there any qualitative researchers here using the discourse graph for coding? I am conducting an interpretative phenomenological analysis on psychotherapists who use psychedelics in their private life: Using the discourse graph and query search to exhaust my data. I could do that quite easily, using the questions from the semi structured interview as QUE's, data from the QUE's as EVD to support claims for the themes.*

**5.2.3 Using Discourse Graphs to Augment Research Training.** Related to this structuring of ongoing research, we observed the potential of the extension to **help beginning researchers learn the process of doing research and synthesis**. Indeed, the potential of discourse graphing to do this was the impetus for launching the field study. In the initial demo of the MVP system in August 2021 to a small set of power users in the Roam community to get a sense of the potential interest and value of the extension, LM, a senior PhD student who was teaching a dissertation writing in Roam in an online course said<sup>26</sup>:

*I see this as a teaching tool. so many people come into a degree, with no idea how to even begin researching yes, so it could be sold as a, this is not a tool to use as you research, but this is a tool to teach you how to research.*

That call led to an initial demo of the extension to a set of students the next day by LK, another senior PhD student who also taught an online course on academic workflows in Roam. The positive initial feedback from the students and the demo spurred a decision to deploy it for the next cohort of his course.

Separately, we also saw evidence of this "learning effect" from other sources of data from the deployment. For example, CV, an independent sociologist, described how the extension enabled her to think differently about how to organize her thoughts<sup>27</sup>:

*Discourse Graph has not only helped me as a tool, but it has made me conceive the process of organizing thought in a different way. It has helped me collapse different related pieces of information around one question/goal.*

Similarly, SD, a cognitive science PhD student, said<sup>28</sup>:

*I liked the way it made me \*think\* about the literature, in terms of Questions/Answers*

### 5.3 Augmenting Collaborative Research with Discourse Graphs

**5.3.1 Opening Vignette.** In the Fall of 2023, approximately 1 year into using the discourse graph to structure ongoing primary research, MA shared how he had used the discourse graph to work with a PhD student to synthesize and summarize a series of results from experiments. The process began with a series of queries over the results discourse nodes associated with their ongoing questions, experiments, and hypotheses for the project. After "dumping" these nodes onto a Playground canvas, MA and the student then used the canvas and nodes to formulate their conclusions for a core research question, structured in terms of a cluster of result nodes that would become the basis of a later compound figure in a publication (see Figure 18 C and D). Reflecting on the process, MA noted <sup>29</sup>:

*I did a 3-hour synthesis session with...(a PhD student) yesterday. I had him convert his results into result nodes (which he had been resisting) and we used a query to drop those nodes + related discourse nodes onto a canvas, to compose into what ended up being two figures. He commented a couple of times how the process called out where he thought he was keeping everything in his head but there were actually parts he had confused or forgot about or duplicated. Especially for the purpose of introducing and communicating the findings to someone else.*

<sup>26</sup>Field note 2021-08-11

<sup>27</sup>Usage survey CV

<sup>28</sup>Usage survey SD

<sup>29</sup>Discord message 2023-11-25

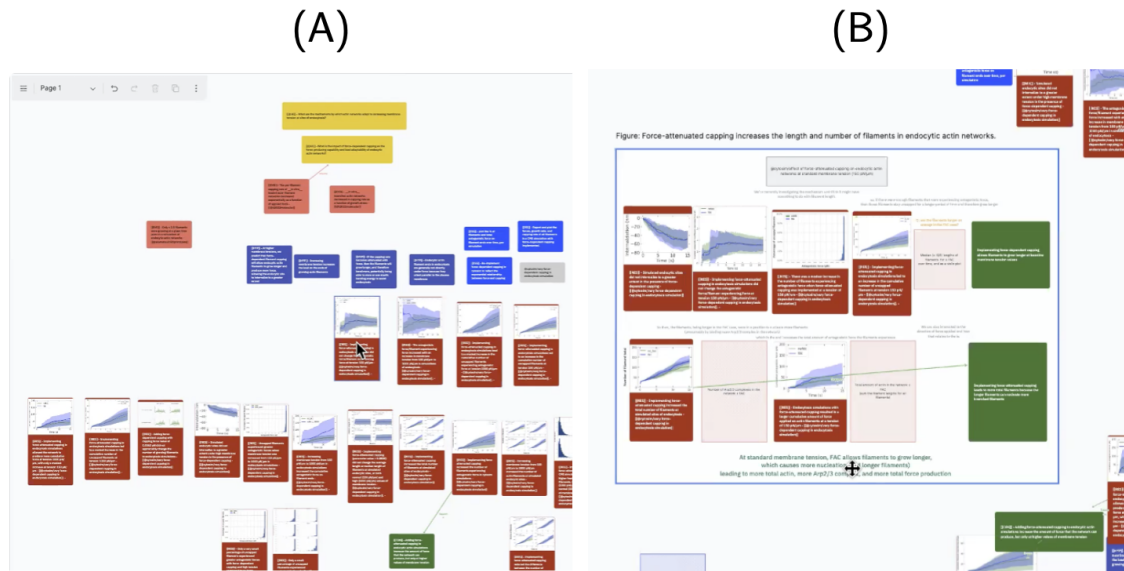


Fig. 15. Snapshots of MA using the discourse graph to collaboratively synthesize experimental results with a PhD student to formulate conclusions for a core research question. MA and the student visually cluster result nodes, shown in maroon here (A), into the beginnings of a "compound figure" with results that substantiate shared conclusions, suitable for inclusion in a publication, shown in dark green here (B).

**5.3.2 Discourse Graphs as Shared Context.** Collaborative usage of the discourse graph — as illustrated by the preceding vignette — was primarily directly observed in the first and fourth author's labs. In the first author's lab, the first author developed a practice of routinely **"bringing in" discourse nodes in collaborative 1-on-1 and group meetings to provide shared context, spur ideation, and leave pointers for followup work**. Figure 16A shows one example of the first author referencing a network analysis result encapsulated in an EVD node in a group meeting to assist with the development of a study plan; the EVD node was used as a quick access point to retrieve an example of the semantic network analysis — including concrete screenshots and methodological details that were attached to the EVD node — that was being proposed for the study<sup>30</sup>. In this specific example, the EVD note was created in March 2022, and referenced in the meeting more than a year later, in June 2023. In a note reflecting on this practice, the first author described its impact on collaborative work like this<sup>31</sup>:

*really tough to kind of... push things through a straw (speaking), and really ineffective and inefficient to try to push through writing. can get alignment through shared sketching. but what about all that context? ...d/graphing is a way to structure your environment to make resources easier to access...bringing it to the table instead of a mere mention*

Students and collaborators were also able to **draw on the growing discourse graph in the lab to scaffold their thinking and writing**. Figure 16B shows an example of how AS, a master's student in the lab, was able to draw on the lab's discourse graph of claims about the role of analogies in creative work, and barriers to benefiting from cross-domain analogies, to scaffold his outlining of his thinking for his thesis: the CLM nodes provided efficient access points to key sources and empirical results to set the context for his problematization, and allowed him to focus on the more

<sup>30</sup>Page sKd3L\_WST in first author's lab Roam graph

<sup>31</sup>Field note 2022-05-23

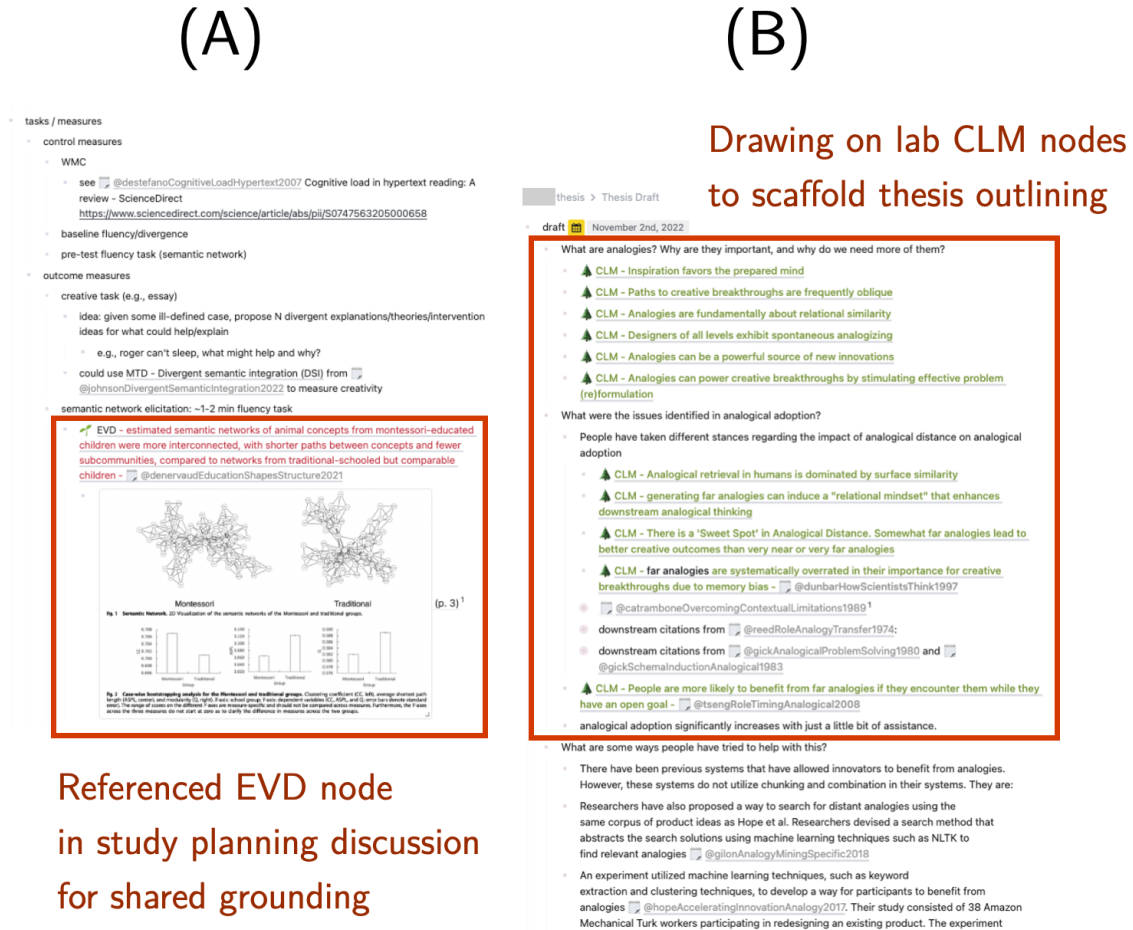


Fig. 16. Snapshots of discourse graphs augmenting collaborative work in the first author's lab: (A) referencing a specific result as an example of a specific type of analysis, to reached shared grounding for planning studies with student advisees, and (B) a master's student scaffolding his outlining of his thesis introduction by building on the lab's library of claims for the formative steps of his argument, allowing him to focus on exploring and expanding the literature review closest to the unique contribution of his thesis.

immediate literature context of his thesis contributions<sup>32</sup>. RH and BL, with whom the first author is collaborating on a research project, also built on existing CLM nodes to propose new hypotheses<sup>33</sup> and also connect analysis of data to existing CLMs<sup>34</sup>.

Similar examples of discourse graphs augmenting collaboration were observed in MA's lab. For example, MA has developed meeting templates that rely on **structured queries over discourse nodes to help him "load context" before meetings** (Figure 17A): the questions, claims, evidence, experiments, and results that are associated with a

<sup>32</sup>Page a2sXaJP4u from first author's lab Roam graph

<sup>33</sup>SS hypertext notebook

<sup>34</sup>Email 2023-11-18

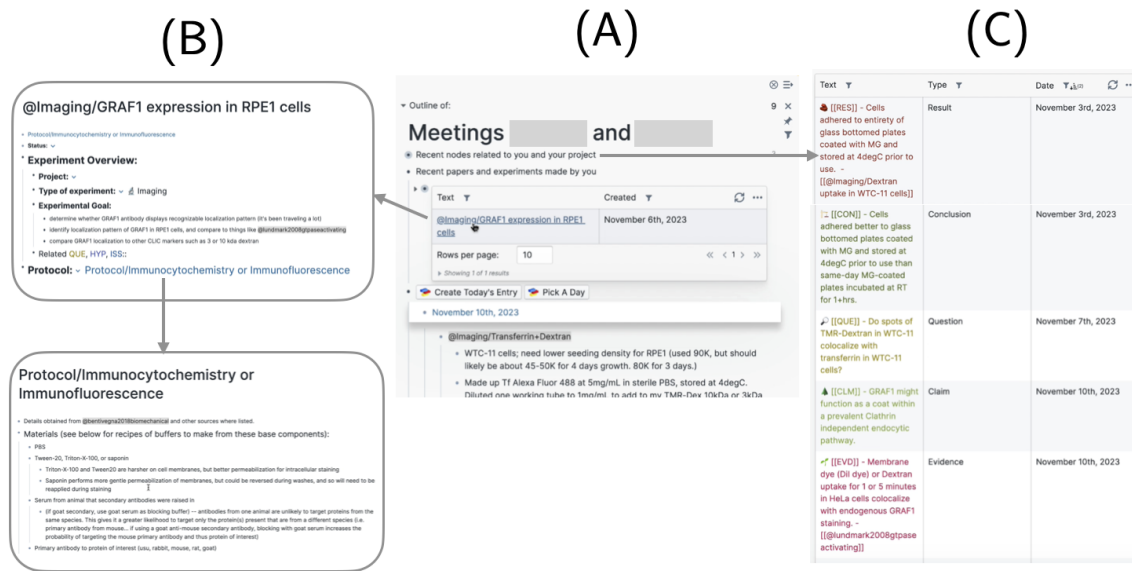


Fig. 17. Snapshot of discourse graph usage enabling shared context in collaborative research meetings: MA uses meeting templates with structured queries over discourse nodes to "load context" before meetings (A). The associated questions, claims, evidence, experiments in the structured queries (C), as well as linked experiment nodes provide efficient access points for relevant details, such as experiment plans and protocols (B), enhancing contextualization.

given meeting (Figure 17C) provide context and efficient access points for needed details, such as experiments and protocols associated with a specific result (Figure 17B)<sup>35</sup>:

*we have this running query about what types of discourse nodes they or we created, which are the scans for discourse nodes created by them, or that I created in this page, or that were created during our group meetings page under heading corresponding to their name. And it's sort of a running scan for the last two weeks, something like that. So that gives me that sort of continual instant query gives me a lot of useful information that I can look into...And so for example, if I say like, actually, I don't know enough about this experiment to comment on it, I can go click and like look at what has been done so far, like what protocol did they use, etc, etc. So all of a sudden I have enough context loaded, enough context loaded to be useful in the meeting.*

**5.3.3 Discourse Graphs as Coordination Structure.** Elaborating more on observations of lab practices for structuring collaboration around discourse graphs, MA's group meetings often **use discourse graphs to provide common language and context to structure collaborative work (moving forward)**. One concrete example was describing and summarizing a weekly update in terms of hypotheses and results, and then creating an issue node to structure needed follow-up work, which was subsequently claimed by an undergraduate researcher, while retaining links to the context of questions and hypotheses (see Figure 18A and B):

*my favorite example that happened recently is that I, you know, here's like a future issue for later. [student's name] started describing her project, and I converted it. I created an issue. Highlights, converts to an issue, and then you get like all this like information like who might contribute, and then somebody can claim the issue. And if they do that, then it gets converted into an experiment that they're running. And so the like the time lag between me creating and this issue and [the student] claiming it was like 30 seconds, which I think was great.*

<sup>35</sup>Screen recording 2023-11-17; Page UA6e6XHVq from MA's lab Roam graph

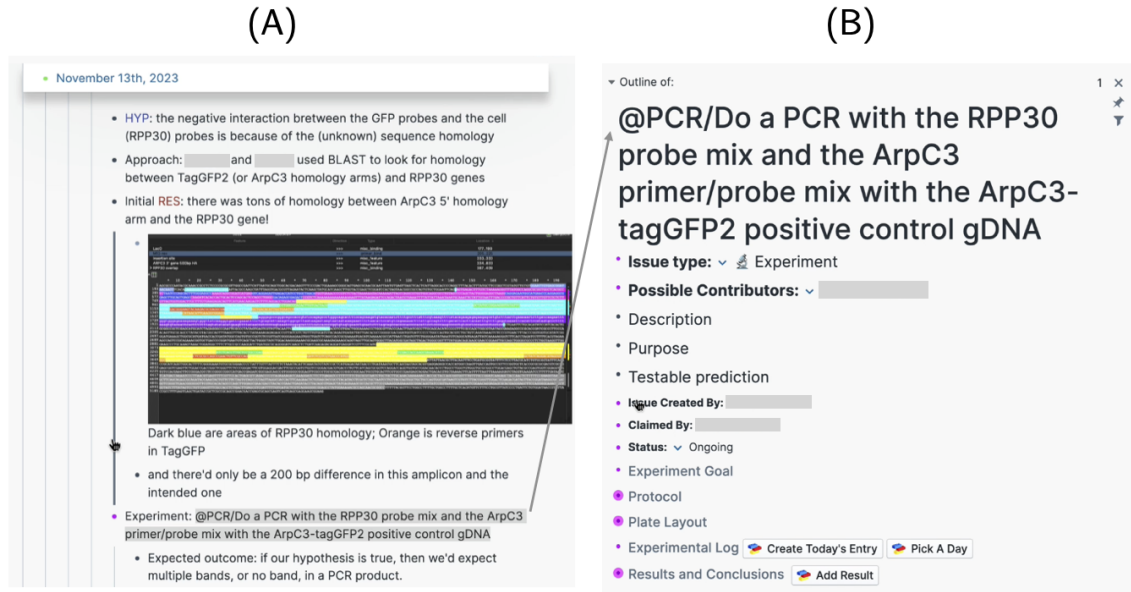


Fig. 18. Snapshots of discourse graphs providing common language and context to structure collaborative work: weekly updates are summarized in terms of hypotheses and results (A), with issue nodes structuring follow-up tasks to test new hypotheses (B).

This structuring and focusing effect of discourse graphs for collaborative research was mentioned by a few other users, including enhancing collaboration of PhD students with their advisors<sup>36</sup>, as well as in a biomedical R&D lab<sup>37</sup>. Together, these data points provide an initial proof of concept that discourse nodes can be useful to someone other than the author, albeit within collaborations with a high degree of shared context, and often in synchronous settings.

## 6 EVALUATION II: SOCIOTECHNICAL PRECONDITIONS FOR A DISCOURSE-CENTRIC INFRASTRUCTURE

In this section of our reflections, we describe how the local transformations of scientific practice described in §5 were enabled by technical approaches for incremental formalization of lab and reading notes into discourse nodes and edges (§6.1), as well as local extensions of the discourse grammar to cover local nuances of practice, including structuring primary research, and supporting synthesis work beyond empirical science (§6.2). We then describe evidence that these discourse-enabled lab notebooks can also be implemented in a range of technical substrates (§6.3), pointing the way to a future where a diverse set of discourse-enabled lab notebooks can be networked in a peer-to-peer fashion using a discourse graph model as a common data exchange protocol.

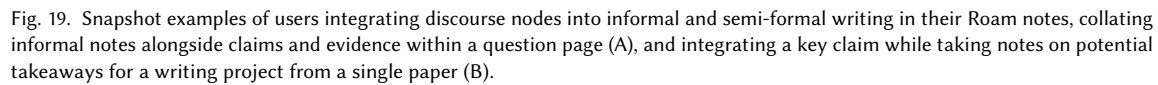
### 6.1 The Possibilities and Limits of Incremental Formalization for Authoring Discourse Nodes and Relations

In our design work on DR1, we hypothesized that discourse graphing can be integrated into everyday scholarly work if it feels as easy as, and is integrated into, existing work like annotating and outlining. Our field data yielded partial

<sup>36</sup>Usage survey QQ; Field note 2022-04-27

<sup>37</sup>Field note 2022-02-08





A range of data points suggest that **users were able to use the node menu to integrate authoring of discourse nodes into their synthesis work**. First, of the four users who shared a csv export of their discourse graph with us in the usage survey, we saw that these users created an average of 1.4k nodes, within the first year of usage (see Figure 27 in Appendix C). Additionally, from the self-report data from the usage survey on how important each feature was to their workflow, 60% of users who (at the time) identified as active users reported that the node menu – which enabled the create-node-as-annotation authoring experience – was core to their workflow. We also saw frequent examples of integration of discourse nodes into semi-formal or informal text, indicating the intended use case of incremental formalization of ideas into discourse nodes where appropriate (as opposed to forcing all authoring to be only discourse nodes; see Figure 19<sup>38</sup>).

<sup>40</sup>Data exports from MA lab 2023-11-04; Field note 2023-11-04

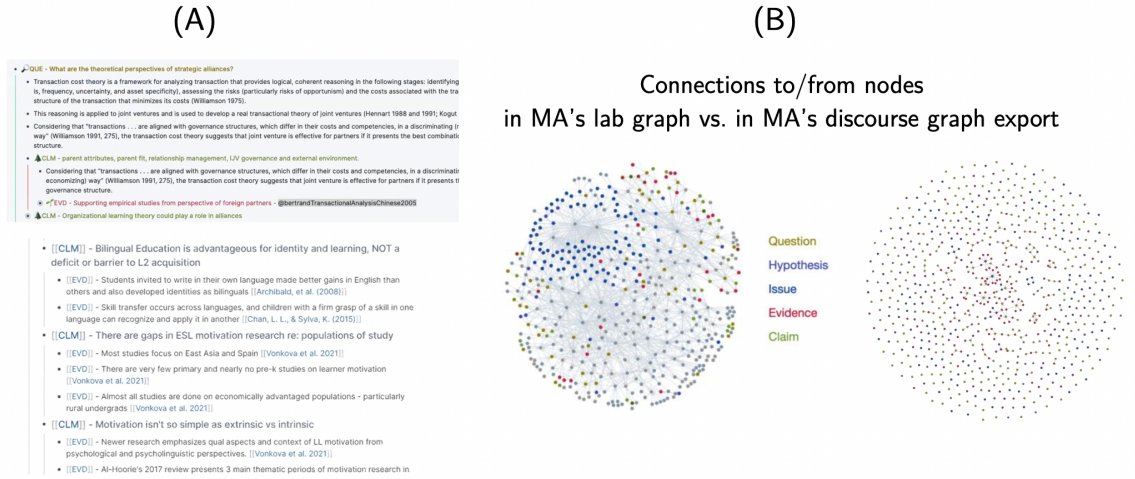


Fig. 20. Snapshot examples of users relating discourse nodes in implicit or informal ways, such as indentation, rather than explicit discourse relations such as "SupportedBy" (A), and an example of the contrast between the richness and density of interconnections between nodes in MA's lab graph (as defined by these less formalized connections; here shown in a screenshot of graph overview visualization in their Roam graph) and the relative sparsity of discourse relations in the discourse graph export from the same lab graph (B).

This lack of explicit discourse relations was less of a limiting factor than we anticipated: users were still able to use the native "bi-directional links" in the Roam graph to navigate and retrieve valuable information from their discourse nodes, as documented in §5.1. CS, a library and information science PhD student, illustrated how this was possible. She was using the extension to collate key takeaways from across literatures on second-language learning, brokerage, and immigrant acculturation, but was not able to share a usable export of the discourse graph for the usage survey because there were essentially no discourse relations. She explained, though, in a followup call, how she was still able to get value out of her discourse nodes<sup>41</sup>:

*"I know that I haven't done it properly in this case, but it doesn't matter, because I can still effectively click on this and see them all, and I can see what ones they're attached to. So it doesn't matter to me that I can't necessarily see, because I can still see TESOL claim evidence. I have my pop. Pop. The things I need to know, plus where it came from. So I don't care that I did it wrong."*

## 6.2 The Importance of Extending and Personalizing the Discourse Grammar

As noted in §4, we anticipated some extension of the discourse grammar, chiefly to cover ways of expressing discourse relations between the base set of questions, claims, and evidence that we may have missed in our initial design. We did not, however, anticipate the **extensive expansion of the base discourse grammar by our users, along with intensive use of the grammar editor**. To orient with some basic statistics from the usage survey, 40% of users self-reported the grammar editor as core to their workflow, with an additional 40% reporting at least occasional usage. Less strictly quantifiable was the frequent discussions of extending/personalizing the discourse grammar on our main Discord channel, and in open office hours and other forums; as a quantitative snapshot, a search for the term "grammar" in our main Discord channel showed 57 individual messages referencing the term "grammar"; many of these messages

<sup>41</sup>Zoom call 2022-09-29



	Node names	Description	Example usage
Structuring primary research	HYP (hypothesis)	HYP and CON are the lab's original claims; HYPotheses spark specific ISSue nodes, which yield RES that substantiate CONclusions	
	CON (conclusion) ISS (issue) RES (result)		
Finer-grained claims	CON (conclusion) PROP (proposition) SCLM (source claims)	CONclusions are higher-level claims that follow from / are composed by sequences of atomic PROPositions; SCLM (source claims) are claims that have only been asserted by a source, not yet linked to empirical evidence	
Bridging to entity graphs	TRM (term) THE (theory)	TRMs are specific defined constructs that are referenced/used in discourse nodes; THEories are named entities that collate claims and constructs and can also be referenced in discourse nodes	
Extending beyond empirical science	PTN (pattern) ART (artifact)	Design knowledge operationalized as design PTNs that are instantiated and test(able) in specific ARTs; can participate in claims and evidence	
	HCON (hypothesized constraint) CON (constraint) HSOL (hypothesized solution) SOL (solution)		

Fig. 21. Example snapshots of key user-driven expansions to the base discourse grammar in terms of new discourse nodes, spanning functions such as structuring primary research, finer-grained claims and argumentation, bridging to entity graphs, and expanding beyond empirical science.

were in the context of conversations or led to threads discussing how to define or extend the grammar. Qualitatively, one of the core users, AP, created his own video describing his grammar expansion for legal research<sup>42</sup>, which was mentioned by several other users as a helpful resource<sup>43</sup>.

The primary way that users extended the discourse grammar was by **defining new types of nodes**. Figure 21 summarizes some key new types of nodes we observed in usage<sup>44</sup>. Some nodes — such as HYPotheses, or RESults — were created to support primary research (described in more detail in MA's usage vignette in §??, as well as in §5.2. Others were created to enable finer-grained claims and argumentation, such as distinguishing between higher-level CONclusions and more atomic PROPositions. TRM (term) nodes were created to bridge the discourse nodes to more defined entities and constructs.

Some nodes were also created to extend beyond empirical science. For example, DSV is a venture capital firm that focuses on investing in companies that translate frontier science into products with high impact on grand challenges such as regenerative agriculture, computation-assisted collective intelligence, climate change, and curative therapeutics. They extended the discourse grammar to enable explicit tracking of *opportunities*, *constraints*, and *solutions*, and their grounding in scientific evidence nodes. This enabled them to reason much more carefully in the process of

<sup>42</sup>AP video walkthrough of discourse graph set up

<sup>43</sup>Open office hours call 2022-04-04

<sup>44</sup>Usage survey MA; Open office hours call with MP 2022-04-20

scoping opportunity areas, including assessment of the necessity and sufficiency of particular venture areas (based on interrelations between constraints and solutions) for grand challenges alongside their level of readiness based on scientific evidence (based on relationships between constraints and solutions and evidence; see Figure 22). In a document describing their motivation for developing their usage of the extension, they wrote<sup>45</sup>:

*[W]e need systems to help us make structured claims and arguments about the possibilities of applied science. We want to be able to answer questions like, "What are the known constraints, both technical and commercial, on achieving societal outcomes?", "What specific pieces of knowledge currently exist related to those constraints?", "How might we uncover counterintuitive or non-obvious intervention points?", "What kinds of expertise are likely to interact effectively with these pieces of knowledge, and who has this expertise?", and "How can we stimulate combinatorial innovation - getting people across domains to speak the same language and making knowledge easy to integrate?"*

This grammar extension was developed with the first author's assistance on a number of user calls, and inspired improvements to the DiscourseGraph extension grammar editor (see §6.2 for more details). While they found value in the way that the extension enabled them to structure their thinking, they have since moved on to use their enhanced discourse grammar in a less formal way to structure their thinking and conversations, and their new experiments with a language-model-powered scoping assistant, which also operates on opportunities, constraints, solutions, and evidence (along with connections and chaining between these node types) as primitives. EM, the team member who had led the implementation of the discourse grammar shared in a semi-structured interview in Fall 2023 that a major reason that they left was the frictions of the Roam software itself, including a lack of affordances for structuring collaborative work, such as version control, comments, and granular permissions; yet, the enduring impact of the extension on their work has been, in her words, *"more of a cultural change than a procedural change"*<sup>46</sup>.

AP also extended the discourse grammar to cover *conclusion* nodes, similar to MA, and also specific relations such as *substantiates*, to enable integration of synthesis into his primary research work in international law (see Figure 23)<sup>47</sup>.

To enable this expansion of the grammar, a substantial portion of our development effort during the deploy period was focused on enhancements to the grammar editor. In addition to small quality-of-life improvements like adding color distinctions between node types, and preserving the layout of nodes in the relation pattern editor, one major design enhancement is worth highlighting here to illustrate the nature of users' expansion of the grammar: **we added the ability to "clone" or "copy-paste" relation patterns** (see Figure 26 in Appendix B). The main motivation for this feature request from users was to ease the cost of defining new discourse node types that were variants or subtypes of existing node types.

For example, MVD, in a direct message on Twitter, described a desire to create subtypes of claims (alluding to the nuances he proposed, as described above) that inherited relation patterns of the base "claim" node<sup>48</sup>:

*What I'd really like is to be able to define subtype of relations (i.e. different kinds of claims). BUT that each subtype automatically inherits the properties of its parent type. i.e. you don't have to add every new claim-relation when you make a new subtype of claim.*

MA expressed a similar request in a Discord message, to assist him in setting up the expansion of the grammar to cover primary research discourse nodes, such as RES (result)<sup>49</sup>:

*I was messing around with getting the grammar to work with RES, didnt explore it thoroughly, wondering if you have thoughts about that (e.g. being able to inherit the grammar from EVD).*

<sup>45</sup>Document 2022-09-22

<sup>46</sup>Zoom call 2023-10-03

<sup>47</sup>AP video walkthrough of discourse graph set up

<sup>48</sup>Twitter direct message 2021-11-22

<sup>49</sup>Discord message 2022-01-08

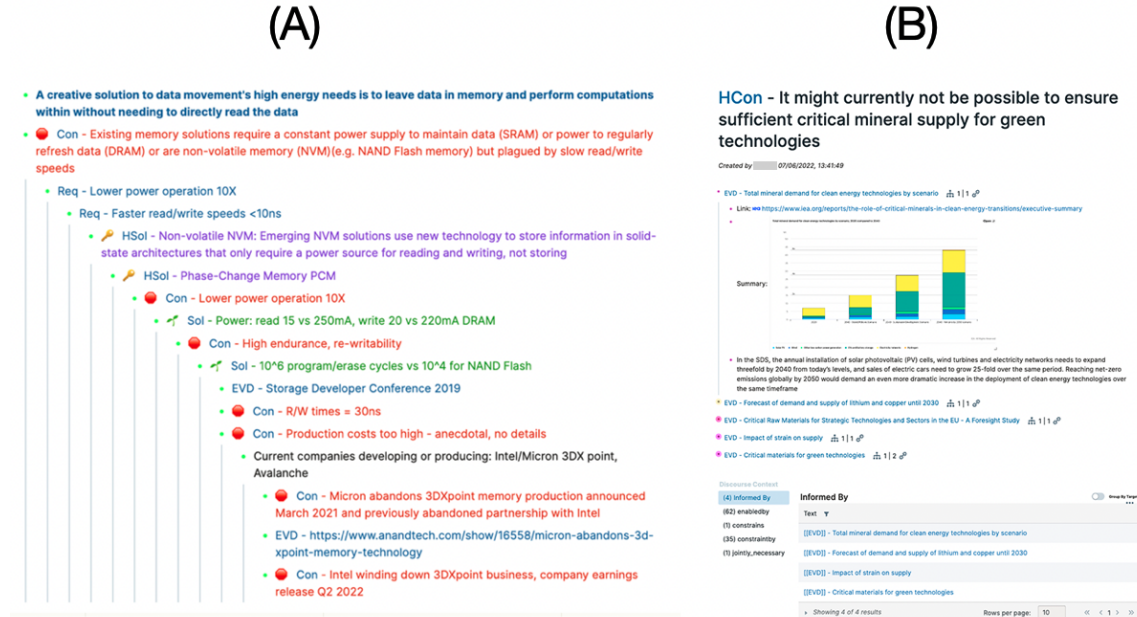


Fig. 22. Snapshots of DSV's usage of the DiscourseGraph extension to scope *constraints*, *requirements*, and *solutions* in an opportunity area for potential investment, alongside scientific *evidence*. (A) shows a scoping tree of requirements, constraints, and solutions for the high level grand challenge of energy efficient data systems, and (B) shows the contents of a page describing a hypothesized constraint around the uncertainty of being able to ensure sufficient critical mineral supply for green technologies, which is grounded in references to specific evidence nodes. The page also surfaces (via the Discourse Context component) specific discourse relations of "enabledby", "constrains", and "jointly necessary", as well "informed by" relations to important pieces of scientific evidence.

There was also some **expansion of the discourse graph grammar to cover new types of discourse relations and relation patterns**. For example, AP added new relations such as "*Substantiates*" between Claims and Conclusions, and "*Makes*" between Authors and Claims (to be able to trace claims back to specific authors)<sup>50</sup>. And DSV added numerous relations to express their such as "*Constrains*" (between Constraints and Solutions) "*Enables*" (between Solutions and Constraints, and between Solutions), and "*Jointly Necessary*" (between Solutions and Constraints/Solutions)<sup>51</sup>. The first author's lab also implemented a "*ConsistentWith*" relation to see "sibling" pieces of evidence that support the same claim (expressing a sort of "conceptual replication" relationship, to give a sense of how "robust" a piece of evidence is in the network of evidence)<sup>52</sup>.

These new relations, though, were less common than the new nodes created. It is hard to determine the extent to which this was a lack of desire for more sophisticated and expressive relations, vs. technical difficulties. Given the expressed and observed desire for smoother and less effortful specification of relations described in the previous section (§6.1), as well as some user data on the continued desire for easier ways to express these patterns to the extension's parser to be able to pick them up (including several instances of the first-author assisting users in defining relation patterns on user calls), we infer that user-friendly recognition of discourse relations remained an open problem in the

<sup>50</sup>Field note 2021-12-16

<sup>51</sup>

<sup>52</sup>Zoom call 2022-07-22

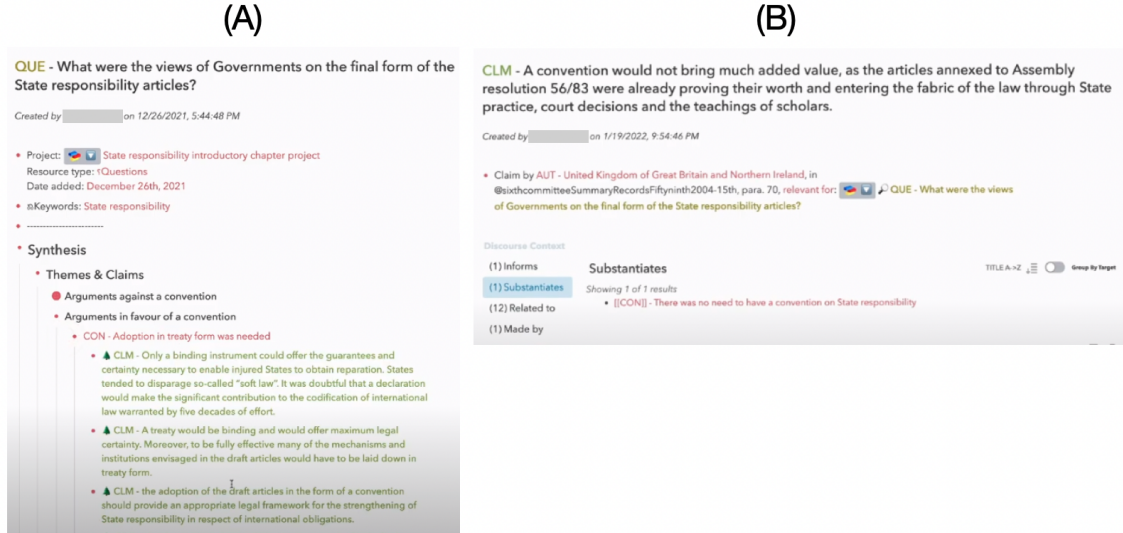


Fig. 23. Snapshots of AP's usage of the DiscourseGraph extension to map out positions and issues for an international law research project. (A) shows a question page that was used to track and synthesize various *claims* and *conclusions* about the views of Governments on the final form of State responsibility articles, and (B) shows a specific page for the claim "A convention would not bring much added value, as the articles annexed to Assembly resolution 56/83 were already proving their worth and entering the fabric of the law through State practice, court decisions and the teachings of scholars". The page tracks the source of the claim (the *Author* being the UK, documented in the @sixthcommittee SummaryRecordsFifty ninth2004-15th source, as well as the claim's discourse relations to other discourse nodes that are highlighted in the Discourse Context component, such as its *substantiates* relation to the *conclusion* that "There was no need to have a convention on State responsibility"

extension. AP's note about the burden of specifying relation patterns in the editor shed some light on this (here his mention of the "playground" refers to the visual editor we provided for specifying relation patterns)<sup>53</sup>:

*It would be great to have a quick way of editing existing relationships, other than returning to the playground. I on occasion "tweak" my existing relationships. For example, Sometimes I want to change a from "child" to "descendant" so as to pick relationships between nodes nested further down the stack. Going back to the extension page and tweaking can take time and can disrupt the flow. Once a relationship is set up, it would be great to have some sort of abbreviated view of the components of the relationship, each of which could be tweaked. In fact, this latter idea could be the basis for eventually abandoning the playground metaphor. While I like the playground, especially since the visual metaphor is useful, it still suffers from the problem that it loses the layout each time you go out and then comeback (at least it only keeps the layout if you go out and return quickly. It is a real pain point to have to reorganize the relationships each time in the visual editor. Either this has to be fixed, or the playground might need to be replaced with another method.*

### 6.3 Transferring the Discourse Graph System across Diverse Technical Settings

One barrier to observation of collaborative usage of discourse graphs was the fact that many scholars with whom our users collaborated already had their own established workflows. This is not surprising, given what we know about the diversity and idiosyncrasy of tooling for scholarly work [15, 90]. To elaborate on this further, we observed several instances of users requesting a "port" of the DiscourseGraph extension to different tools they were more familiar with or preferred to work in, for various reasons. For example, an early-career faculty member, explained in their usage

<sup>53</sup>Usage survey AP

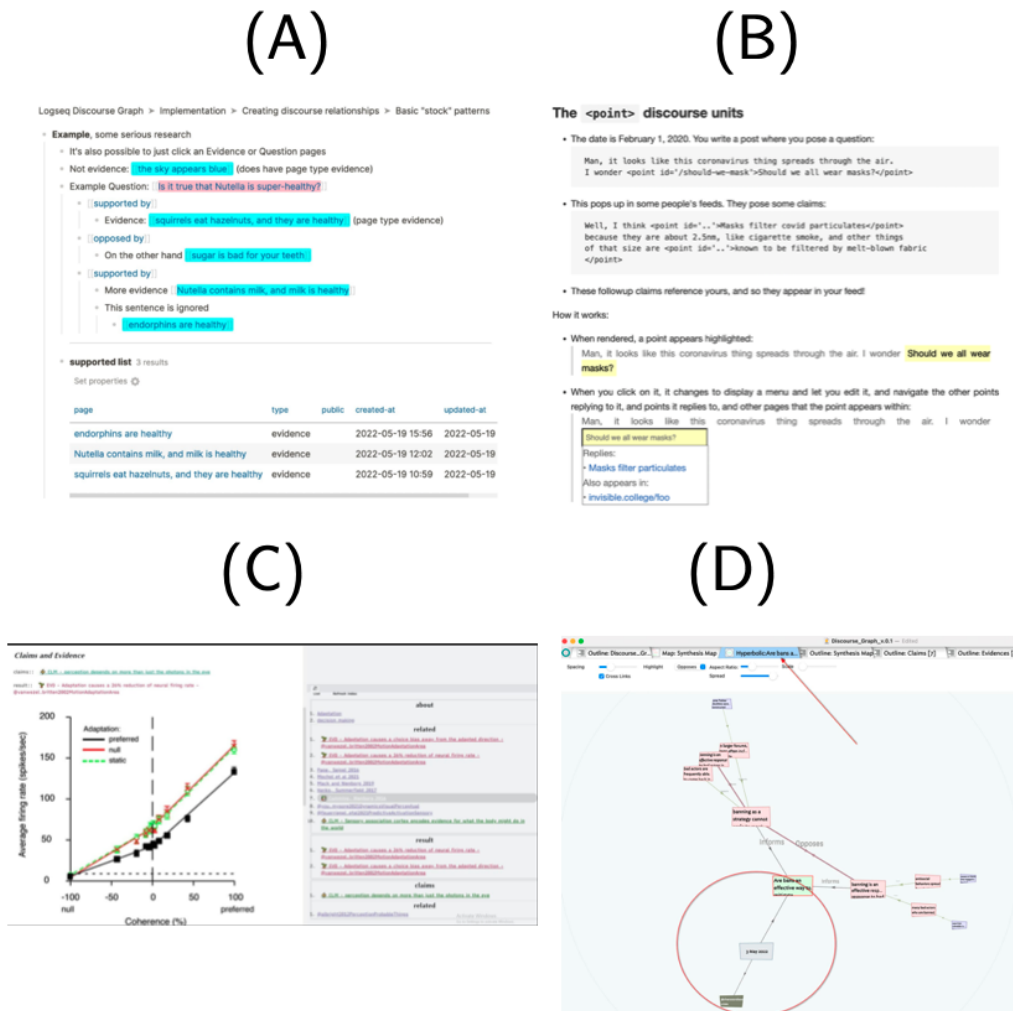


Fig. 24. Snapshots of observed instances of user-driven implementation of discourse graphs in a similar outliner hypertext notebook (Logseq) (A), an HTML markup library for web publishing (B), a markdown-based hypertext notebook (Obsidian) (C), and a visual hypertext notebook (Tinderbox) (D).

survey how they preferred to do their synthesis work in Obsidian, a markdown-based hypertext notebook, and was attempting to (manually) adapt the Discourse Graph workflow in their work there<sup>54</sup>:

*I haven't used discourse graph extension as much as I might just because I've been using Roam a lot less recently. This is because I do my synthesis work in Obsidian because it is easier to transition to writing there. But if there were an extension for Obsidian, I would 100% use it because my solution is kind of a hack.*

<sup>54</sup>Usage survey MD

We were heartened, therefore, to observe **organic spread of the Discourse Graph practice and software beyond Roam**: through our involvement in the "tools for thought" community on forums and community chats, social media, and calls with users, we observed implementations of Discourse Graphs in Obsidian<sup>55</sup> (see Figure 24C), Tinderbox (a longstanding visual hypertext notebook; see Figure 24D), Tana (an outliner hypertext notebook similar to Roam but with more powerful in-built primitives for applying schemas to notes<sup>56</sup>). Notably, users from the Logseq community organized an \$8,000 bounty (funded in part by a grant from one of their institutions) for a full port of the DiscourseGraph extension to Logseq; over the Spring through Fall of 2022, those users, and the developers they recruited, involved the first author in several working meetings to develop specifications for a minimum viable prototype port<sup>57</sup>; Figure 24B shows the progress of the prototype, which was released to users for testing in December 2023<sup>58</sup>. MA also sparked efforts to prototype a discourse graph authoring markup in a web publishing framework (see Figure 24B)<sup>59</sup>.

## 7 DISCUSSION

In this paper, we explored the HCI problem of how to grow new knowledge infrastructures optimized for sharing, reusing, and synthesizing knowledge. Over a 2.5-year span of participatory Research through Design with a distributed community of researchers using hypertext notebooks, we sketched out a design vision for how to integrate the seeds of a discourse-centric infrastructure into local scientific practices, and the collaborative and institutional structures of research. Our field deployment highlighted the following key aspects of this vision:

- **Intrinsic benefits** of discourse graphing for improved synthesis (§5.1), transformed primary research and research training (§5.2), and augmented collaboration (§5.3)
- **Sociotechnical preconditions** for successful integration, including smooth(er) incremental formalization [96] of notes into discourse nodes and edges (§6.1), and local personalization of discourse graph grammars (§6.2)
- **Technical means of networking** these local discourse graphs into a larger infrastructure by integrating discourse graphs into a diversity of notetaking tools, while retaining a shared data model (§6.3)

We can map our results to Edwards et al's [34] historical model of infrastructure development to understand how these results contribute steps towards growing an infrastructure for synthesis. From the perspective of this model, the discourse-enhanced lab notebooks are **local systems** that augment synthesis. The authoring of discourse graphs in an interoperable format, and evidence of their use in collaborative research work, are seeds of distributed **networking** of these local synthesis systems. And the organic spread of the discourse-enhanced lab notebook pattern from our original technical setting of Roam Research to other technical settings, and from our original empirical science setting to other substantive domains, constitute seeds of **technology transfer** across different contexts. We are especially excited about this point, since infrastructure studies have observed that technology transfer is often a key catalyst for growing local systems into infrastructure [34]. For these reasons, we claim that our results go beyond illuminating a design vision: they also contribute concrete **steps towards an infrastructure for synthesis**.

We close by discussing the open HCI research problems our design vision illuminates for continuing the growth of a discourse-centric research synthesis infrastructure, as well as broader implications of our work for HCI.

<sup>55</sup>Field note 2022-01-13; Field note 2022-04-13

<sup>56</sup>Field note 2022-08-20

<sup>57</sup>Field note 2022-04-26

<sup>58</sup>Discord message 2023-12-09

<sup>59</sup>Field note 2022-02-11

### 7.1 How HCI can Contribute to Continued Growth of an Infrastructure for Synthesis

As we have argued, the effective local systems for synthesis and transfer across technical contexts constitute a trajectory towards infrastructure. We see the next phase of work being fruitfully focused on *networking* local discourse graphs: what work remains to continue this trajectory, and what can HCI contribute?

First, as described in §6.1, we observed limitations in terms of being able to smoothly author discourse edges. This is in part because the mechanisms by which our system was able to recognize patterns of discourse was limited to fairly coarse primitives of the hypertext medium (of indentation or linking, for example): to enable more specific semantics, such as differentiating between support and oppose relations, users had to manually specify and execute the relevant patterns. We wonder whether users might be able to more smoothly recognize latent discourse edges in their existing work with systems that integrate AI approaches for transforming unstructured (text) data into more structured formats. There is already a history of work like this in the field of (semi-)automated evidence synthesis [52, 69, 106]: we are curious whether the apparent leaps in capacity from transformer models and large pretrained models [11] might make AI assistance more immediately useful. This could be an interesting problem to connect with research on mixed-initiative intelligent user interfaces [47, 50]. This problem could also be framed as one of empowering users — often with little to no existing programming expertise — to “program” their notes, thereby connecting with research on end-user programming [77].

Second, we also observed some “non-atomic” authoring of discourse nodes. For example, some nodes used undefined abbreviations, or were written with unresolved references. Evidence nodes were also often “empty” placeholders: they lacked direct descriptions or links to methodological details or contextualizing snippets from papers. These nodes were still useful for individuals and within synchronous, tight-knit collaborative work contexts, and provided some ready-made on-ramps for conceptually significant “legitimate peripheral participation” [62] for beginning research assistants (as described in §5.2 and Fig. 12B). However, CSCW research on the knowledge sharing [2] suggests that additional labor is likely needed to make them useful beyond these tight-knit settings. Here, again, we think there may be interesting research directions around mixed-initiative interfaces that integrate advances in argument mining [92], claim generation [115], or scientific document information extraction [53] to assist with rewriting and contextualization of discourse graphs for sharing.

Third, as described in §6.2, we saw the core data model of questions, claims, and evidence, being frequently locally extended, while still preserving core properties across the local variations. For example, the core “claim” node type bridged between a past-focused variant of synthesis work (similar to systematic reviews) and the emerging practice of micropublishing ongoing primary research. And the core “evidence” node type also bridged between empirical science and design innovation. This dynamic of minimal shared global semantics paired with local variation the core notion of *boundary objects* [103] that facilitate coordination across social worlds and can be developed into standards for infrastructure formation [65]. We are curious to see more direct tests of whether discourse graphs could act as boundary objects in actual practice. For example, what would happen if we enabled peer-to-peer exchange of local discourse graphs across, say, evidence synthesis groups like Cochrane, biomedical research labs, and biotech innovation companies and investing firms? Would these groups be able to coordinate and share information usefully via the common nodes of claims and evidence, while extending and connecting them to local extensions, such as hypotheses and issues (for primary research), or constraints and solutions (for innovation)? This research could illuminate the next set of (socio)technical problems that HCI researchers might be able to tackle to enable federation or networking across local discourse graph systems. From one perspective, this could be framed as a purely technical problem, and approach



it similarly to research on ontology alignment and integration. But given the deeply sociotechnical nature of negotiation and alignment that constitutes standards development [34], HCI research might make important contributions by drawing on existing design patterns for supporting collective deliberation [57] and infrastructuring [63, 82].

## 7.2 Broader implications for HCI

Our work also has broader implications of for HCI beyond metascience. For instance, the successful application of incremental formalization as a design pattern in the context of research synthesis further validates the power of that design pattern for building interactive systems that support creative knowledge work. The power of the discourse graph formalism can also be understood as convergent validation of the power of domain-specific languages (DSL) [38] for augmenting individual and collaborative intelligence, such as the IBIS system [26, 61], Fischer’s work on the Envisionment and Discovery Collaboratory (EDC) [7, 8], Liddo et al’s work on applying similar discourse-centric models to structure and augment contested collective deliberation [28, 66], and models of cross-domain analogical retrieval based on an underlying schema of matching problems and mechanisms [55] to name just a few. We think that more exploration of varieties of domain-specific languages might broaden the settings of creative work that HCI is able to augment. These domain-specific languages might also form the basis of more effective “shared representations” [47] that enable more transparent, controllable and effective development of mixed-initiative systems, or more user-controlled retrieval-augmented generation (RAG) workflows [40] for creativity support.

To close, we reiterate that we found key concepts from sociotechnical and critical studies of infrastructure and knowledge sharing — such as the frame of growing infrastructure, or boundary objects — to be immensely helpful for guiding our design work. These concepts went beyond descriptions of what should *not* be built or what cannot work (e.g., naive repository models for knowledge sharing, as described in [2]): they helped us envision what *could* be built: in this case, a new infrastructure for synthesis grown from a new installed based of enhanced synthesis and research notetaking systems. We see our work therefore as contributing an additional example of how HCI research might fruitfully bridge the “great divide” between system building and social analysis [16]. We believe the integration of critique and systems building is increasingly necessary as HCI seeks to respond to pressing sociotechnical and infrastructure-scale problems, such as misinformation, and privacy and security amid the proliferation of generative AI.

## ACKNOWLEDGMENTS

This work was made possible in part by grants from Protocol Labs Research to the first, second, and fourth authors. We are deeply indebted also to the members of the RoamResearch and larger tools for thought community, and to Wayne Lutters, Jodi Schneider, Katrina Fenlon, Evan Miyazono, Karola Kirsanow, Silvia diBessa, SJ Klein, and Robert Haisfield for thoughtful comments on the conceptualization of this work.

## REFERENCES

- [1] Margunn Aanestad, Miria Grisot, Ole Hanseth, and Polyxeni Vassilakopoulou. 2017. Information infrastructures and the challenge of the installed base. In *Information Infrastructures within European Health Care*, Margunn Aanestad, Miria Grisot, Ole Hanseth, and Polyxeni Vassilakopoulou (Eds.). Springer International Publishing, Cham, 25–33. <https://doi.org/10.1007/978-3-319-51020-0>
- [2] 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, 4-6 (Aug. 2013), 531–573. <https://doi.org/10.1007/s10606-013-9192-8>
- [3] Adrienne Alton-Lee. 1998. A Troubleshooter’s Checklist for Prospective Authors Derived from Reviewers’ Critical Feedback. *Teaching and Teacher Education* 14, 8 (1998), 887–90.
- [4] Mats Alvesson and Jörgen Sandberg. 2011. Generating research questions through problematization. *Academy of management review* 36, 2 (2011), 247–271. Publisher: Academy of Management Briarcliff Manor, NY.



- [5] Paul André, Haoqi Zhang, Juho Kim, Lydia Chilton, Steven P. Dow, and Robert C. Miller. 2013. Community clustering: Leveraging an academic crowd to form coherent conference sessions. In *First AAAI Conference on Human Computation and Crowdsourcing*.
- [6] anujacabraal. 2012. Why use NVivo for your literature review? <https://anujacabraal.wordpress.com/2012/08/01/why-use-nvivo-for-your-literature-review/>
- [7] Ernesto Arias, Hal Eden, Gerhard Fischer, Andrew Gorman, and Eric Scharff. 2000. Transcending the individual human mind—creating shared understanding through collaborative design. *ACM Transactions on Computer-Human Interaction* 7, 1 (March 2000), 84–113. <https://doi.org/10.1145/344949.345015>
- [8] Ernesto G. Arias, Hal Eden, and Gerhard Fischer. 2015. The Envisionment and Discovery Collaboratory (EDC): Explorations in Human-Centered Informatics. *Synthesis Lectures on Human-Centered Informatics* 8, 5 (Oct. 2015), i–216. <https://doi.org/10.2200/S00670ED1V01Y201509HCI032> Publisher: Morgan & Claypool Publishers.
- [9] Sheetal Bhurke, Andrew Cook, Anna Tallant, Amanda Young, Elaine Williams, and James Raftery. 2015. Using systematic reviews to inform NIHR HTA trial planning and design: a retrospective cohort. *BMC Medical Research Methodology* 15, 1 (Dec. 2015), 108. <https://doi.org/10.1186/s12874-015-0102-2> 00000.
- [10] Catherine Blake and Wanda Pratt. 2006. Collaborative information synthesis I: A model of information behaviors of scientists in medicine and public health. *Journal of the American Society for Information Science and Technology* 57, 13 (2006), 1740–1749. <https://doi.org/10.1002/asi.20487> 00097\_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/asi.20487>
- [11] Rishi Bommasani, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein, Jeannette Bohg, Antoine Bosselut, Emma Brunskill, Erik Brynjolfsson, Shyamal Buch, Dallas Card, Rodrigo Castellon, Niladri Chatterji, Annie Chen, Kathleen Creel, Jared Quincy Davis, Dora Demszky, Chris Donahue, Moussa Doumbouya, Esin Durmus, Stefano Ermon, John Etchemendy, Kawin Ethayarajh, Li Fei-Fei, Chelsea Finn, Trevor Gale, Lauren Gillespie, Karan Goel, Noah Goodman, Shelby Grossman, Neel Guha, Tatsunori Hashimoto, Peter Henderson, John Hewitt, Daniel E. Ho, Jenny Hong, Kyle Hsu, Jing Huang, Thomas Icard, Saahil Jain, Dan Jurafsky, Pratyusha Kalluri, Siddharth Karamcheti, Geoff Keeling, Fereshte Khani, Omar Khattab, Pang Wei Koh, Mark Krass, Ranjay Krishna, Rohith Kuditipudi, Ananya Kumar, Faisal Ladhak, Mina Lee, Tony Lee, Jure Leskovec, Isabelle Levent, Xiang Lisa Li, Xuechen Li, Tengyu Ma, Ali Malik, Christopher D. Manning, Suvir Mirchandani, Eric Mitchell, Zanele Munyikwa, Suraj Nair, Avanika Narayan, Deepak Narayanan, Ben Newman, Allen Nie, Juan Carlos Niebles, Hamed Nilforoshan, Julian Nyarko, Giray Ogut, Laurel Orr, Isabel Papadimitriou, Joon Sung Park, Chris Piech, Eva Portelance, Christopher Potts, Aditi Raghunathan, Rob Reich, Hongyu Ren, Frieda Rong, Yusuf Roohani, Camilo Ruiz, Jack Ryan, Christopher Ré, Dorsa Sadigh, Shiori Sagawa, Keshav Santhanam, Andy Shih, Krishnan Srinivasan, Alex Tamkin, Rohan Taori, Armin W. Thomas, Florian Tramèr, Rose E. Wang, William Wang, Bohan Wu, Jiajun Wu, Yuhuai Wu, Sang Michael Xie, Michihiro Yasunaga, Jiaxuan You, Matei Zaharia, Michael Zhang, Tianyi Zhang, Xikun Zhang, Yuhui Zhang, Lucia Zheng, Kaitlyn Zhou, and Percy Liang. 2021. On the Opportunities and Risks of Foundation Models. *arXiv:2108.07258 [cs]* (Aug. 2021). <http://arxiv.org/abs/2108.07258> 00021 arXiv: 2108.07258.
- [12] David N. Boote and Penny Beile. 2005. Scholars Before Researchers: On the Centrality of the Dissertation Literature Review in Research Preparation. *Educational Researcher* 34, 6 (Aug. 2005), 3–15. <https://doi.org/10.3102/0013189X034006003>
- [13] Rohit Borah, Andrew W. Brown, Patrice L. Capers, and Kathryn A. Kaiser. 2017. Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry. *BMJ Open* 7, 2 (Feb. 2017), e012545. <https://doi.org/10.1136/bmjopen-2016-012545> Publisher: British Medical Journal Publishing Group Section: Health informatics.
- [14] Christine L Borgman, Paul N Edwards, Steven J Jackson, Melissa K Chalmers, Geoffrey C Bowker, David Ribes, Matt Burton, and Scout Calvert. 2013. Knowledge Infrastructures: Intellectual Frameworks and Research Challenges. (2013).
- [15] Jeroen Bosman and Bianca Kramer. 2016. Innovations in scholarly communication - data of the global 2015-2016 survey. <https://doi.org/10.5281/zenodo.49583> 00014 type: dataset.
- [16] Geoffrey Bowker, Susan Leigh Star, Les Gasser, and William Turner. 2014. *Social Science, Technical Systems, and Cooperative Work: Beyond the Great Divide*. Psychology Press. Google-Books-ID: XSaPAwAAQBAJ.
- [17] Geoffrey C. Bowker. 2000. Biodiversity Datadiversity. *Social Studies of Science* 30, 5 (Oct. 2000), 643–683. <https://doi.org/10.1177/030631200030005001> Publisher: SAGE Publications Ltd.
- [18] Matthew H. Brush, Kent Shefchek, and Melissa Haendel. 2016. SEPIO: A semantic model for the integration and analysis of scientific evidence. *CEUR Workshop Proceedings* 1747 (2016). <https://ohsu.pure.elsevier.com/en/publications/seprio-a-semantic-model-for-the-integration-and-analysis-of-scient>
- [19] Cristina-Iulia Bucur, Tobias Kuhn, Davide Ceolin, and Jacco van Ossenberg. 2022. Nanopublication-Based Semantic Publishing and Reviewing: A Field Study with Formalization Papers. *arXiv:2203.01608 [cs]* (March 2022). <http://arxiv.org/abs/2203.01608> arXiv: 2203.01608.
- [20] Joel Chan. 2020. Knowledge synthesis: A conceptual model and practical guide. *Open and Sustainable Innovation Systems (OASIS) Lab* (Dec. 2020). <https://oasislab.pubpub.org/pub/54t0y9mk/release/2> 00000 Publisher: PubPub.
- [21] Eshwar Chandrasekharan, Shagun Jhaver, Amy Bruckman, and Eric Gilbert. 2022. Quarantined! Examining the Effects of a Community-Wide Moderation Intervention on Reddit. *ACM Transactions on Computer-Human Interaction* 29, 4 (Aug. 2022), 1–26. <https://doi.org/10.1145/3490499>
- [22] Lydia B. Chilton, Juho Kim, Paul André, Felicia Cordeiro, James A. Landay, Daniel S. Weld, Steven P. Dow, Robert C. Miller, and Haoqi Zhang. 2014. Frenzy: Collaborative Data Organization for Creating Conference Sessions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 1255–1264. <https://doi.org/10.1145/2556288.2557375>

- [23] Paolo Ciccarese, Elizabeth Wu, Gwen Wong, Marco Ocana, June Kinoshita, Alan Ruttenberg, and Tim Clark. 2008. The SWAN Biomedical Discourse Ontology. *Journal of Biomedical Informatics* 41, 5 (Oct. 2008), 739–751. <https://doi.org/10.1016/j.jbi.2008.04.010>
- [24] Tim Clark, Paolo N. Ciccarese, and Carole A. Goble. 2014. Micropublications: A Semantic Model for Claims, Evidence, Arguments and Annotations in Biomedical Communications. *Journal of Biomedical Semantics* 5 (July 2014), 28. <https://doi.org/10.1186/2041-1480-5-28>
- [25] Tim Clark and June Kinoshita. 2007. Alzforum and SWAN: the present and future of scientific web communities. *Briefings in Bioinformatics* 8, 3 (May 2007), 163–171. <https://doi.org/10.1093/bib/bbm012.00074>.
- [26] Jeff Conklin. 1987. Hypertext: An Introduction and Survey. *Computer* 20, 9 (1987), 17–41.
- [27] Éric Dagiral and Ashveen Peerbaye. 2016. Making Knowledge in Boundary Infrastructures: Inside and Beyond a Database for Rare Diseases. *Science & Technology Studies* 29, 2 (May 2016), 44–61. <https://doi.org/10.23987/sts.55920> Number: 2.
- [28] Anna De Liddo and Rosa Strube. 2021. Understanding Failures and Potentials of Argumentation Tools for Public Deliberation. In *Proceedings of the 10th International Conference on Communities & Technologies - Wicked Problems in the Age of Tech (C&T '21)*. Association for Computing Machinery, New York, NY, USA, 75–88. <https://doi.org/10.1145/3461564.3461584>
- [29] Hélène de Ribaupierre and Gilles Falquet. 2017. Extracting Discourse Elements and Annotating Scientific Documents Using the SciAnnotDoc Model: A Use Case in Gender Documents. *International Journal on Digital Libraries* (Aug. 2017), 1–16. <https://doi.org/10.1007/s00799-017-0227-5>
- [30] Anita de Waard. 2010. From Proteins to Fairytales: Directions in Semantic Publishing. *IEEE Intelligent Systems* (2010).
- [31] Anita de Waard, Simon Buckingham Shum, Annamaria Carusi, Jack Park, Matthias Samwald, and Agnes Sándor. 2009. Hypotheses, Evidence and Relationships: The Hyper Approach for Representing Scientific Knowledge Claims. In *Proceedings of the 8th International Semantic Web Conference, Workshop on Semantic Web Applications in Scientific Discourse*. 12.
- [32] Paul N. Edwards. 2010. *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*. MIT Press. Google-Books-ID: K9\_LsJBCqWMC.
- [33] Paul N Edwards, Geoffrey C Bowker, Steven J Jackson, and Robin Williams. 2009. Introduction: An Agenda for Infrastructure Studies. *Journal of the Association for Information Systems* 10, 5 (2009), 6.
- [34] Paul N Edwards, Steven J Jackson, Geoffrey C Bowker, and Cory P Knobel. 2007. *Understanding infrastructure: Dynamics, tensions, and design*. Technical Report.
- [35] Paul N. Edwards, Matthew S. Mayernik, Archer L. Batcheller, Geoffrey C. Bowker, and Christine L. Borgman. 2011. Science friction: Data, metadata, and collaboration. *Social Studies of Science* 41, 5 (Oct. 2011), 667–690. <https://doi.org/10.1177/0306312711413314> Publisher: SAGE Publications Ltd.
- [36] David Ellis. 1993. Modeling the Information-Seeking Patterns of Academic Researchers: A Grounded Theory Approach. *The Library Quarterly* 63, 4 (Oct. 1993), 469–486. <https://doi.org/10.1086/602622>
- [37] Ann-Margret Ervin. 2008. Motivating authors to update systematic reviews: practical strategies from a behavioural science perspective. *Paediatric and perinatal epidemiology* 22, 0 1 (Jan. 2008), 33–37. <https://doi.org/10.1111/j.1365-3016.2007.00910.x>
- [38] Gerhard Fischer. 1994. Turning breakdowns into opportunities for creativity. *Knowledge-Based Systems* 7, 4 (Dec. 1994), 221–232. [https://doi.org/10.1016/0950-7051\(94\)90033-7](https://doi.org/10.1016/0950-7051(94)90033-7) 00000.
- [39] Padhraig S. Fleming, Jadbinder Seehra, Argy Polychronopoulou, Zbys Fedorowicz, and Nikolaos Pandis. 2013. Cochrane and non-Cochrane systematic reviews in leading orthodontic journals: a quality paradigm? *European Journal of Orthodontics* 35, 2 (April 2013), 244–248. <https://doi.org/10.1093/ejo/cjs016> Publisher: Oxford Academic.
- [40] Yunfan Gao, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yi Dai, Jiawei Sun, Qianyu Guo, Meng Wang, and Haofen Wang. 2024. Retrieval-Augmented Generation for Large Language Models: A Survey. <https://doi.org/10.48550/arXiv.2312.10997> arXiv:2312.10997 [cs].
- [41] William Gaver. 2014. Science and design: The implications of different forms of accountability. In *Ways of knowing in HCI*, Judith S. Olson and Wendy A. Kellogg (Eds.). Springer, New York NY, 143–165.
- [42] Darcy Haag Granello. 2001. Promoting Cognitive Complexity in Graduate Written Work: Using Bloom’s Taxonomy as a Pedagogical Tool to Improve Literature Reviews. *Counselor Education and Supervision* 40, 4 (June 2001), 292–307. <https://doi.org/10.1002/j.1556-6978.2001.tb01261.x>
- [43] Malachi Griffith, Nicholas C. Spies, Kilannin Krysiak, Joshua F. McMichael, Adam C. Coffman, Arpad M. Danos, Benjamin J. Ainscough, Cody A. Ramirez, Damian T. Rieke, Lynzey Kujan, Erica K. Barnell, Alex H. Wagner, Zachary L. Skidmore, Amber Wollam, Connor J. Liu, Martin R. Jones, Rachel L. Bilski, Robert Lesurf, Yan-Yang Feng, Nakul M. Shah, Melika Bonakdar, Lee Trani, Matthew Matlock, Avinash Ramu, Katie M. Campbell, Gregory C. Spies, Aaron P. Graubert, Karthik Gangavarapu, James M. Eldred, David E. Larson, Jason R. Walker, Benjamin M. Good, Chunlei Wu, Andrew I. Su, Rodrigo Dienstmann, Adam A. Margolin, David Tamborero, Nuria Lopez-Bigas, Steven J. M. Jones, Ron Bose, David H. Spencer, Lukas D. Wartman, Richard K. Wilson, Elaine R. Mardis, and Obi L. Griffith. 2017. CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. *Nature Genetics* 49, 2 (Feb. 2017), 170–174. <https://doi.org/10.1038/ng.3774> 00239 Number: 2 Publisher: Nature Publishing Group.
- [44] Paul Groth, Andrew Gibson, and Jan Velterop. 2010. The anatomy of a nanopublication. *Information Services & Use* 30, 1-2 (Jan. 2010), 51–56. <https://doi.org/10.3233/ISU-2010-0613>
- [45] Tudor Groza, Knud Möller, Siegfried Handschuh, Diana Trif, and Stefan Decker. 2007. SALT: Weaving the Claim Web. In *The Semantic Web (Lecture Notes in Computer Science)*, Karl Aberer, Key-Sun Choi, Natasha Noy, Dean Allemang, Kyung-Il Lee, Lyndon Nixon, Jennifer Golbeck, Peter Mika, Diana Maynard, Riichiro Mizoguchi, Guus Schreiber, and Philippe Cudré-Mauroux (Eds.). Springer Berlin Heidelberg, 197–210.
- [46] Ole Hanseth and Kristin Braa. 1998. Technology as Traitor: Emergent SAP Infrastructure in a Global Organization. *ICIS 1998 Proceedings* (Dec. 1998). <https://aisel.aisnet.org/icis1998/17>

- [47] Jeffrey Heer. 2019. Agency plus Automation: Designing Artificial Intelligence into Interactive Systems. *Proceedings of the National Academy of Sciences* 116, 6 (Feb. 2019), 1844–1850. <https://doi.org/10.1073/pnas.1807184115>
- [48] Linh Hoang and Jodi Schneider. 2018. Opportunities for Computer Support for Systematic Reviewing - A Gap Analysis. In *Transforming Digital Worlds (Lecture Notes in Computer Science)*, Gobinda Chowdhury, Julie McLeod, Val Gillet, and Peter Willett (Eds.). Springer International Publishing, 367–377.
- [49] Allyson Holbrook, Sid Bourke, Terence Lovat, and Kerry Dally. 2004. Investigating PhD thesis examination reports. *International Journal of Educational Research* 41, 2 (Jan. 2004), 98–120. <https://doi.org/10.1016/j.ijer.2005.04.008>
- [50] Eric Horvitz. 1999. Principles of Mixed-Initiative User Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '99)*. ACM, New York, NY, USA, 159–166. <https://doi.org/10.1145/302979.303030>
- [51] Thomas Parke Hughes. 1983. *Networks of Power: Electrification in Western Society, 1880-1930*. JHU Press. 05700 Google-Books-ID: g07Q9M4agp4C.
- [52] Siddhartha R. Jonnalagadda, Pawan Goyal, and Mark D. Huffman. 2015. Automating data extraction in systematic reviews: a systematic review. *Systematic Reviews* 4, 1 (June 2015), 78. <https://doi.org/10.1186/s13643-015-0066-7>
- [53] Halil Kilicoglu. 2017. Biomedical Text Mining for Research Rigor and Integrity: Tasks, Challenges, Directions. *Briefings in Bioinformatics* (2017). <https://doi.org/10.1093/bib/bbx057>
- [54] Juho Kim, Phu Tran Nguyen, Sarah Weir, Philip J. Guo, Robert C. Miller, and Krzysztof Z. Gajos. 2014. Crowdsourcing Step-by-step Information Extraction to Enhance Existing How-to Videos. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 4017–4026. <https://doi.org/10.1145/2556288.2556986>
- [55] Aniket Kittur, Lixiu Yu, Tom Hope, Joel Chan, Hila Lifshitz-Assaf, Karni Gilon, Felicia Ng, Robert E. Kraut, and Dafna Shahaf. 2019. Scaling up Analogical Innovation with Crowds and AI. *Proceedings of the National Academy of Sciences* (Jan. 2019), 201807185. <https://doi.org/10.1073/pnas.1807185116>
- [56] Ian A. Knight, Max L. Wilson, David F. Brailsford, and Natasha Milic-Frayling. 2019. Enslaved to the Trapped Data: A Cognitive Work Analysis of Medical Systematic Reviews. In *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval (CHIIR '19)*. ACM, New York, NY, USA, 203–212. <https://doi.org/10.1145/3295750.3298937> event-place: Glasgow, Scotland UK.
- [57] Travis Kriplean, Jonathan Morgan, Deen Freelon, Alan Borning, and Lance Bennett. 2012. Supporting Reflective Public Thought with Considerit. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (CSCW '12)*. ACM, New York, NY, USA, 265–274. <https://doi.org/10.1145/2145204.2145249> 00000.
- [58] Tobias Kuhn. 2009. AceWiki: A Natural and Expressive Semantic Wiki. In *Semantic Web User Interaction (SWUI) CEUR Workshop*. 8. 00092.
- [59] Tobias Kuhn, Paolo Emilio Barbano, Mate Levente Nagy, and Michael Krauthammer. 2013. Broadening the Scope of Nanopublications. In *The Semantic Web: Semantics and Big Data (Lecture Notes in Computer Science)*, Philipp Cimiano, Oscar Corcho, Valentina Presutti, Laura Hollink, and Sebastian Rudolph (Eds.). Springer Berlin Heidelberg, 487–501.
- [60] Tobias Kuhn and Michel Dumontier. 2017. Genuine semantic publishing. *Data Science* 1, 1-2 (Jan. 2017), 139–154. <https://doi.org/10.3233/DS-170010>
- [61] Werner Kunz and Horst WJ Rittel. 1970. *Issues as elements of information systems*. Working Paper 131. Institute of Urban and Regional Development, University of California Berkeley, Berkeley, California, USA.
- [62] Jean Lave and Etienne Wenger. 1991. *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press.
- [63] Christopher A Le Dantec and Carl DiSalvo. 2013. Infrastructuring and the formation of publics in participatory design. *Social Studies of Science* 43, 2 (April 2013), 241–264. <https://doi.org/10.1177/0306312712471581> 00469.
- [64] Charlotte P. Lee, Paul Dourish, and Gloria Mark. 2006. The Human Infrastructure of Cyberinfrastructure. In *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work (CSCW '06)*. ACM, New York, NY, USA, 483–492. <https://doi.org/10.1145/1180875.1180950> event-place: Banff, Alberta, Canada.
- [65] Susan Leigh Star. 2010. This Is Not a Boundary Object: Reflections on the Origin of a Concept , This Is Not a Boundary Object: Reflections on the Origin of a Concept. *Science, Technology, & Human Values* 35, 5 (Sept. 2010), 601–617. <https://doi.org/10.1177/0162243910377624>
- [66] Anna De Liddo, Ágnes Sándor, and Simon Buckingham Shum. 2012. Contested Collective Intelligence: Rationale, Technologies, and a Human-Machine Annotation Study. *Computer Supported Cooperative Work (CSCW)* 21, 4-5 (Oct. 2012), 417–448. <https://doi.org/10.1007/s10606-011-9155-x>
- [67] Barbara E. Lovitts. 2007. *Making the Implicit Explicit: Creating Performance Expectations for the Dissertation*. Stylus Publishing, Sterling, Va.
- [68] Catherine C. Marshall, Frank M. Shipman, and James H. Coombs. 1994. VIKI: spatial hypertext supporting emergent structure. In *Proceedings of the 1994 ACM European conference on Hypermedia technology (EHT '94)*. Association for Computing Machinery, New York, NY, USA, 13–23. <https://doi.org/10.1145/192757.192759>
- [69] Iain J. Marshall, Blair T. Johnson, Zigeng Wang, Sanguthevar Rajasekaran, and Byron C. Wallace. 2020. Semi-Automated evidence synthesis in health psychology: current methods and future prospects. *Health Psychology Review* 14, 1 (Jan. 2020), 145–158. <https://doi.org/10.1080/17437199.2020.1716198> 2020.1716198 Publisher: Routledge \_eprint: <https://doi.org/10.1080/17437199.2020.1716198>
- [70] D. Scott McCrickard. 2012. Making Claims: Knowledge Design, Capture, and Sharing in HCI. *Synthesis Lectures on Human-Centered Informatics* 5, 3 (June 2012), 1–125. <https://doi.org/10.2200/S00423ED1V01Y201205HCI015>
- [71] D. Scott McCrickard, Shahtab Wahid, Stacy M. Branham, and Steve Harrison. 2013. Achieving Both Creativity and Rationale: Reuse in Design with Images and Claims. In *Creativity and Rationale*, John M. Carroll (Ed.). Number 20 in Human-Computer Interaction Series. Springer London, 105–119. [https://doi.org/10.1007/978-1-4471-4111-2\\_6](https://doi.org/10.1007/978-1-4471-4111-2_6)

- [72] Richard McElreath and Paul E. Smaldino. 2015. Replication, Communication, and the Population Dynamics of Scientific Discovery. *PLOS ONE* 10, 8 (Aug. 2015), e0136088. <https://doi.org/10.1371/journal.pone.0136088> Publisher: Public Library of Science.
- [73] Jonathon McPhetres, Nihan Albayrak-Aydemir, Ana Barbosa Mendes, Elvina C. Chow, Patricio Gonzalez-Marquez, Erin Loukras, Annika Maus, Aoife O'Mahony, Christina Pomareda, Maximilian Primbs, Shalaine Sackman, Conor Smithson, and Kirill Volodko. 2020. *A decade of theory as reflected in Psychological Science (2009-2019)*. Technical Report. PsyArXiv. <https://doi.org/10.31234/osf.io/hs5nx> 00000 type: article.
- [74] Matthew Michelson and Katja Reuter. 2019. The significant cost of systematic reviews and meta-analyses: A call for greater involvement of machine learning to assess the promise of clinical trials. *Contemporary Clinical Trials Communications* 16 (Dec. 2019), 100443. <https://doi.org/10.1016/j.conctc.2019.100443>
- [75] John S Morabito and Joel Chan. 2021. Managing Context during Scholarly Knowledge Synthesis: Process Patterns and System Mechanics. In *Creativity and Cognition (C&C '21)*. Association for Computing Machinery, New York, NY, USA, 1–5. <https://doi.org/10.1145/3450741.3465244> 00000.
- [76] Gaia Mosconi, Qinyu Li, Dave Randall, Helena Karasti, Peter Tolmie, Jana Barutzky, Matthias Korn, and Volkmar Pipek. 2019. Three Gaps in Opening Science. *Computer Supported Cooperative Work (CSCW)* 28, 3 (June 2019), 749–789. <https://doi.org/10.1007/s10606-019-09354-z>
- [77] Brad A. Myers, Amy J. Ko, and Margaret M. Burnett. 2006. Invited research overview: end-user programming. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems (CHI EA '06)*. Association for Computing Machinery, New York, NY, USA, 75–80. <https://doi.org/10.1145/1125451.1125472>
- [78] T. H. Nelson. 1965. Complex information processing: a file structure for the complex, the changing and the indeterminate. In *Proceedings of the 1965 20th national conference (ACM '65)*. Association for Computing Machinery, Cleveland, Ohio, USA, 84–100. <https://doi.org/10.1145/800197.806036>
- [79] Kenton O'Hara, Fiona Smith, William Newman, and Abigail Sellen. 1998. Student readers' use of library documents: implications for library technologies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '98)*. ACM Press/Addison-Wesley Publishing Co., Los Angeles, California, USA, 233–240. <https://doi.org/10.1145/274644.274678> 00000.
- [80] Carole L. Palmer, Lauren C. Teffeu, and Carrie M. Pirmann. 2009. *Scholarly Information Practices in the Online Environment: Themes from the Literature and Implications for Library Service Development*. Technical Report. 59 pages. <http://www.oclc.org/programs/publications/reports/2009-02.pdf>
- [81] Anthony Petrosino. 1999. Lead authors of cochrane reviews: Survey results. *Report to the Campbell Collaboration*. Cambridge, MA: University of Pennsylvania (1999).
- [82] Volkmar Pipek and Volker Wulf. 2009. Infrastructuring: Toward an Integrated Perspective on the Design and Use of Information Technology. *Journal of the Association for Information Systems* 10, 5 (May 2009), 447–473. <https://doi.org/10.17705/1jais.00195>
- [83] Xin Qian, Katrina Fenlon, Wayne G. Lutters, and Joel Chan. 2020. Opening Up the Black Box of Scholarly Synthesis: Intermediate Products, Processes, and Tools.. In *Proceedings of ASIST 2020*.
- [84] Allen H. Renear and Carole L. Palmer. 2009. Strategic Reading, Ontologies, and the Future of Scientific Publishing. *Science* 325, 5942 (Aug. 2009), 828–832. <https://doi.org/10.1126/science.1157784> 00000.
- [85] A. H. Renear and C. L. Palmer. 2009. Strategic reading, ontologies, and the future of scientific publishing. *Science* 325, 5942 (2009), 828.
- [86] David Ribes and Charlotte P. Lee. 2010. Sociotechnical Studies of Cyberinfrastructure and e-Research: Current Themes and Future Trajectories. *Computer Supported Cooperative Work (CSCW)* 19, 3 (Aug. 2010), 231–244. <https://doi.org/10.1007/s10606-010-9120-0>
- [87] Betsy Rolland and Charlotte P. Lee. 2013. Beyond trust and reliability: reusing data in collaborative cancer epidemiology research. In *Proceedings of the 2013 conference on Computer supported cooperative work (CSCW '13)*. Association for Computing Machinery, New York, NY, USA, 435–444. <https://doi.org/10.1145/2441776.2441826>
- [88] Daniel M. Russell, Mark J. Stefik, Peter Pirolli, and Stuart K. Card. 1993. The Cost Structure of Sensemaking. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems (CHI '93)*. ACM, New York, NY, USA, 269–276. <https://doi.org/10.1145/169059.169209>
- [89] Elizabeth B.-N. Sanders and Pieter Jan Stappers. 2008. Co-creation and the new landscapes of design. *CoDesign* 4, 1 (March 2008), 5–18. <https://doi.org/10.1080/15710880701875068>
- [90] Steve Sawyer, Elizabeth Kaziunas, and Carsten Øesterlund. 2012. Social Scientists and Cyberinfrastructure: Insights from a Document Perspective. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (CSCW '12)*. ACM, New York, NY, USA, 931–934. <https://doi.org/10.1145/2145204.2145342>
- [91] Anne M. Scheel, Leonid Tiokhin, Peder M. Isager, and Daniël Lakens. 2020. Why Hypothesis Testers Should Spend Less Time Testing Hypotheses. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science* (Dec. 2020), 1745691620966795. <https://doi.org/10.1177/1745691620966795> 00008.
- [92] Jodi Schneider, Tudor Groza, and Alexandre Passant. 2013. A review of argumentation for the social semantic web. *Semantic Web* 4, 2 (2013), 159–218. Publisher: IOS Press.
- [93] Thomas Schwandt. 1994. Constructivist, Interpretivist Approaches to Human Inquiry. In *Handbook of Qualitative Research*. SAGE, Thousand Oaks, California.
- [94] Donald A. Schön. 1983. *The reflective practitioner: How professionals think in action*. Basic books.
- [95] Joseph Seering, Robert Kraut, and Laura Dabbish. 2017. Shaping Pro and Anti-Social Behavior on Twitch Through Moderation and Example-Setting. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. ACM, New York, NY, USA, 111–125. <https://doi.org/10.1145/2998181.2998277>

- [96] Frank M. Shipman and Catherine C. Marshall. 1999. Formality Considered Harmful: Experiences, Emerging Themes, and Directions on the Use of Formal Representations in Interactive Systems. *Computer Supported Cooperative Work (CSCW)* 8, 4 (Dec. 1999), 333–352. <https://doi.org/10.1023/A:1008716330212>
- [97] Frank M. Shipman and Raymond McCall. 1994. Supporting knowledge-base evolution with incremental formalization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '94)*. Association for Computing Machinery, Boston, Massachusetts, USA, 285–291. <https://doi.org/10.1145/191666.191768> 00000.
- [98] Kaveh G. Shojania, Margaret Sampson, Mohammed T. Ansari, Jun Ji, Steve Doucette, and David Moher. 2007. How Quickly Do Systematic Reviews Go Out of Date? A Survival Analysis. *Annals of Internal Medicine* 147, 4 (Aug. 2007), 224. <https://doi.org/10.7326/0003-4819-147-4-200708210-00179>
- [99] Simon Buckingham Shum, Enrico Motta, and John Domingue. 2000. ScholOnto: An Ontology-Based Digital Library Server for Research Documents and Discourse. *International Journal on Digital Libraries* 3, 3 (Oct. 2000), 237–248. <https://doi.org/10.1007/s007990000034>
- [100] Simon J. Buckingham Shum, Victoria Uren, Gangmin Li, Bertrand Sereno, and Clara Mancini. 2006. Modeling Naturalistic Argumentation in Research Literatures: Representation and Interaction Design Issues. *International Journal of Intelligent Systems* 22, 1 (Nov. 2006), 17–47. <https://doi.org/10.1002/int.20188>
- [101] Pao Siangliulue, Joel Chan, Steven P. Dow, and Krzysztof Z. Gajos. 2016. IdeaHound: Improving Large-Scale Collaborative Ideation with Crowd-Powered Real-Time Semantic Modeling. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. ACM, New York, NY, USA, 609–624. <https://doi.org/10.1145/2984511.2984578>
- [102] Christina Silver. 2020. The crux of literature reviewing: structuring critical appraisals and using CAQDAS-packages. <https://www.qdaservices.co.uk/post/the-crux-of-literature-reviewing-structuring-critical-appraisals-and-using-caqdas-packages> Library Catalog: [www.qdaservices.co.uk](http://www.qdaservices.co.uk)
- [103] Susan Leigh 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>
- [104] Susan Leigh 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. <https://pubsonline.informs.org/doi/abs/10.1287/isre.7.1.111> 03070 tex.publisher: INFORMS.
- [105] Kenneth Strike and George Posner. 1983. Types of synthesis and their criteria. In *Knowledge Structure and Use*, S Ward and L Reed (Eds.). Temple University Press, Philadelphia, PA.
- [106] James Thomas, Anna Noel-Storr, Iain Marshall, Byron Wallace, Steven McDonald, Chris Mavergames, Paul Glasziou, Ian Shemilt, Anneliese Synnot, Tari Turner, Julian Elliott, Thomas Agoritsas, John Hilton, Caroline Perron, Elie Akl, Rebecca Hodder, Charlotte Petridge, Lauren Albrecht, Tanya Horsley, Joanne Platt, Rebecca Armstrong, Phi Hung Nguyen, Robert Plovnick, Anneliese Arno, Noah Ivers, Gail Quinn, Agnes Au, Renea Johnston, Gabriel Rada, Matthew Bagg, Arwel Jones, Philippe Ravaud, Catherine Boden, Lara Kahale, Bernt Richter, Isabelle Boisvert, Homa Keshavarz, Rebecca Ryan, Linn Brandt, Stephanie A. Kolakowsky-Hayner, Dina Salama, Alexandra Brazinova, Sumanth Kumbargere Nagraj, Georgia Salanti, Rachel Buchbinder, Toby Lasserson, Lina Santaguida, Chris Champion, Rebecca Lawrence, Nancy Santesso, Jackie Chandler, Zbigniew Les, Holger J. Schünemann, Andreas Charidimou, Stefan Leucht, Ian Shemilt, Roger Chou, Nicola Low, Diana Sherifali, Rachel Churchill, Andrew Maas, Reed Siemieniuk, Maryse C. Cnossen, Harriet MacLehose, Mark Simmonds, Marie-Joelle Cossi, Malcolm Macleod, Nicole Skoetz, Michel Counotte, Iain Marshall, Karla Soares-Weiser, Samantha Craigie, Rachel Marshall, Velandai Srikanth, Philipp Dahm, Nicole Martin, Katrina Sullivan, Alanna Danilkewich, Laura Martínez García, Anneliese Synnot, Kristen Danko, Chris Mavergames, Mark Taylor, Emma Donoghue, Lara J. Maxwell, Kris Thayer, Corinna Dressler, James McAuley, James Thomas, Cathy Egan, Steve McDonald, Roger Tritton, Julian Elliott, Joanne McKenzie, Guy Tsafnat, Sarah A. Elliott, Joerg Meerpohl, Peter Tugwell, Itziar Etzeandia, Bronwen Merner, Alexis Turgeon, Robin Featherstone, Stefania Mondello, Tari Turner, Ruth Foxlee, Richard Morley, Gert van Valkenhoef, Paul Garner, Marcus Munafo, Per Vandvik, Martha Gerrity, Zachary Munn, Byron Wallace, Paul Glasziou, Melissa Murano, Sheila A. Wallace, Sally Green, Kristine Newman, Chris Watts, Jeremy Grimshaw, Robby Nieuwlaat, Laura Weeks, Kurinchi Gurusamy, Adriani Nikolakopoulou, Aaron Weigl, Neal Haddaway, Anna Noel-Storr, George Wells, Lisa Hartling, Annette O’Connor, Wojtek Wiercioch, Jill Hayden, Matthew Page, Luke Wolfenden, Mark Helfand, Manisha Pahwa, Juan José Yepes Nuñez, Julian Higgins, Jordi Pardo Pardo, Jennifer Yost, Sophie Hill, and Leslea Pearson. 2017. Living Systematic Reviews: 2. Combining Human and Machine Effort. *Journal of Clinical Epidemiology* 91 (Nov. 2017), 31–37. <https://doi.org/10.1016/j.jclinepi.2017.08.011>
- [107] Stephen E. Toulmin. 2003. *The Uses of Argument*. Cambridge University Press. 00064 Google-Books-ID: 8UYgegaB1S0C.
- [108] Victoria Uren, Simon Buckingham Shum, Michelle Bachler, and Gangmin Li. 2006. Sensemaking Tools for Understanding Research Literatures: Design, Implementation and User Evaluation. *International Journal of Human-Computer Studies* 64, 5 (May 2006), 420–445. <https://doi.org/10.1016/j.ijhcs.2005.09.004>
- [109] Max Van Kleek, Michael Bernstein, David R. Karger, and mc schraefel. 2007. Gui — Phooey!: The Case for Text Input. In *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology (UIST '07)*. ACM, New York, NY, USA, 193–202. <https://doi.org/10.1145/1294211.1294247>
- [110] Iris van Rooij and Giosuè Baggio. 2021. Theory Before the Test: How to Build High-Verisimilitude Explanatory Theories in Psychological Science. *Perspectives on Psychological Science* (Jan. 2021), 1745691620970604. <https://doi.org/10.1177/1745691620970604> 00029 Publisher: SAGE Publications Inc.
- [111] Lisa M. Vaughn and Farrah Jacquez. 2020. Participatory Research Methods – Choice Points in the Research Process. *Journal of Participatory Research Methods* 1, 1 (July 2020). <https://doi.org/10.35844/001c.13244>
- [112] Sarah Weir, Juho Kim, Krzysztof Z. Gajos, and Robert C. Miller. 2015. Learnersourcing Subgoal Labels for How-to Videos. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15)*. ACM, New York, NY, USA, 405–416. <https://doi.org/10.1145/2675133.2675219>

- [113] Matt Willis, Sarika Sharma, Jaime Snyder, Michelle Brown, Carsten Østerlund, and Steve Sawyer. 2014. Documents and Distributed Scientific Collaboration. In *Proceedings of the Companion Publication of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW Companion '14)*. ACM, New York, NY, USA, 257–260. <https://doi.org/10.1145/2556420.2556491>
- [114] Joost F Wolfswinkel, Elfi Furtmueller, and Celeste P M Wilderom. 2013. Using grounded theory as a method for rigorously reviewing literature. *European Journal of Information Systems* 22, 1 (Jan. 2013), 45–55. <https://doi.org/10.1057/ejis.2011.51> Publisher: Taylor & Francis.
- [115] Dustin Wright, David Wadden, Kyle Lo, Bailey Kuehl, Arman Cohan, Isabelle Augenstein, and Lucy Lu Wang. 2022. Generating Scientific Claims for Zero-Shot Scientific Fact Checking. *arXiv:2203.12990 [cs]* (March 2022). <http://arxiv.org/abs/2203.12990> arXiv: 2203.12990.
- [116] Y. Yamamoto, S. Takada, M. Gross, and K. Nakakoji. 1998. Representational talkback: an approach to support writing as design. In *Proceedings. 3rd Asia Pacific Computer Human Interaction (Cat. No.98EX110)*. 125–131. <https://doi.org/10.1109/APCHI.1998.704176>
- [117] Alyson L. Young and Wayne G. Lutters. 2017. Infrastructuring for Cross-Disciplinary Synthetic Science: Meta-Study Research in Land System Science. *Computer Supported Cooperative Work (CSCW)* 26, 1 (April 2017), 165–203. <https://doi.org/10.1007/s10606-017-9267-z>
- [118] John Zimmerman and Jodi Forlizzi. 2014. Research Through Design in HCI. In *Ways of Knowing in HCI*, Judith S. Olson and Wendy A. Kellogg (Eds.). Springer New York, New York, NY, 167–189. [https://doi.org/10.1007/978-1-4939-0378-8\\_8](https://doi.org/10.1007/978-1-4939-0378-8_8)

## A USAGE SURVEY QUESTIONS

- (1) **What is your occupation?** *Options:*
  - Research - Academic
  - Research - Independent / applied
  - Other
- (2) **Where are you in the world?** Please share city/state and country as you feel comfortable. (*Free text*)
- (3) **What sorts of things do you work on?** If you're a researcher, you can briefly describe your general research area(s) and some examples of problems you work on. If you're not doing research, briefly describe what you do. Think about the length of a Twitter bio or so. (*Free text*)
- (4) If Occupation = "Research - Academic":
  - (a) What is the nature of your academic research position? *Options:*
    - Undergraduate Student Researcher
    - Graduate Student Researcher - Masters
    - Graduate Student Researcher - PhD
    - Postdoctoral researcher
    - Tenure-track faculty - pre-tenure
    - Tenure-track faculty - post-tenure
    - Professional / teaching faculty - junior
    - Professional / teaching faculty - senior
    - Other full-time position (e.g., research scientist)
    - Other
- (5) Are you a current user of the extension? *Options:*
  - Yes
  - Maybe - still exploring if will fit
  - No - tried but didn't stick
- (6) If user = "Yes"
  - (a) **How did you learn about and start using the extension?** Common sources include Roam courses, Twitter, and Slack/Discord. We want to understand the channels through which the extension is spreading: what communities or sub-communities seem to be open to or need something like this. We also want to understand what sorts of support is needed to learn to use the extension. (*Free text*)

- (b) **Why did you start using the extension? What value is it providing to you now?** We want to understand the value proposition of this extension and its concepts for other users like yourself. It would be especially helpful if you are able to share specific examples of the extension's functionality delivering value to you, perhaps in contrast to how things used to be before you started using the extension.
- (c) **Have you have you shared any part of your discourse graph with others? Why and how? What went well / could go better?** We want to understand the value proposition of this extension and its concepts for improving collaborative/distributed synthesis.
- (d) **Which of these features do you use?** If you're unsure of the list of features, you might learn more here: <https://oasis-lab.gitbook.io/roamresearch-discourse-graph-extension/> (*Options for each feature: "Didn't know about this", "Never used", "Use on occasion", "Core part of my workflow"*)
- Node menu (for creating nodes)
  - Discourse context
  - Query drawer (ad-hoc)
  - Discourse overlay
  - Grammar editor (for creating/modifying nodes and relation patterns)
  - Export
  - Playground
  - Node discourse attributes
  - Node index
  - Node templates
  - Query drawer (saved to pages)
- (e) **Is there a screenshot (or set of screenshots) or video walkthrough you are willing to share to illustrate your usage of the extension and what value it provides to you?** (*File upload*)
- (f) **Alternatively is there a link to a recording (Loom/Youtube) you are willing to share to illustrate your usage of the extension and what value it provides to you?**
- (g) **If you shared a screenshot/video, would you be ok with me sharing this with other potential users and in research reports to help others understand the extension?** If you say no, I will use it only for confidential analysis. (*Options: "Yes", "No"*)
- (h) **Would you be willing to share your csv export?** We are interested in understanding how people are using and modifying the grammar. We would like to analyze how many nodes and edges have been created, and how interconnected these are in your graph. If possible, we would also like to understand how the core question/claim/evidence nodes are used and written. If you are concerned about privacy, please feel free to remove the title and/or author names from the nodes csv before uploading here. If you wish to do so, and are not sure how: open the command palette (CMD/CTRL + P) and select "Export Discourse Graph" and then select the Neo4j export type. You will be prompted to save a .zip file, from which you can extract two csv files (<yourgraphname>\_query-results\_<datetime>\_nodes.csv and <yourgraphname>\_query-results\_<datetime>\_relations.csv). Feel free to attach the zip file, or the modified csv files separately. (*File upload*)
- (i) **What are your biggest pain points or wishlist items at the moment with the extension?** We are interested in understanding the main blockers to something like this extension being used by users like yourself, and what we might be able to do to remove those blockers. (*Free text*)

(7) If user = “Maybe”

- (a) **How did you learn about and start exploring the extension?** Common sources include Roam courses, Twitter, and Slack/Discord. We want to understand the channels through which the extension is spreading: what communities or sub-communities seem to be open to or need something like this. We also want to understand what sorts of support is needed to learn to use the extension. (*Free text*)
- (b) **Why did you start exploring the extension? What value do you hope it could provide to you?** We want to understand the potential value proposition of this extension and its concepts for other users like yourself. It would be especially helpful if you are able to share specific examples of how you hope the extension’s functionality would deliver value to you, perhaps in contrast to how things are right now in your workflows and work.
- (c) **Which of these features have you explored?** If you’re unsure of the list of features, you might learn more here: <https://oasis-lab.gitbook.io/roamresearch-discourse-graph-extension/> (*Options for each feature: “Didn’t know about this”, “Never used”, “Use on occasion”, “Core part of my workflow”*)
  - Node menu (for creating nodes)
  - Discourse context
  - Query drawer (ad-hoc)
  - Discourse overlay
  - Grammar editor (for creating/modifying nodes and relation patterns)
  - Export
  - Playground
  - Node discourse attributes
  - Node index
  - Node templates
  - Query drawer (saved to pages)
- (d) **What are your biggest pain points or wishlist items at the moment with the extension?** We are interested in understanding the main blockers to something like this extension being used by users like yourself, and what we might be able to do to remove those blockers. (*Free text*)

(8) If user = “No”

- (a) **How did you learn about and start exploring the extension?** Common sources include Roam courses, Twitter, and Slack/Discord. We want to understand the channels through which the extension is spreading: what communities or sub-communities seem to be open to or need something like this. We also want to understand what sorts of support is needed to learn to use the extension. (*Free text*)
- (b) **Why did you start exploring the extension? What value did you hope it could provide to you?** We want to understand the potential value proposition of this extension and its concepts for other users like yourself. It would be especially helpful if you are able to share specific examples of how you hoped the extension’s functionality would have delivered value to you, perhaps in contrast to how things are right now in your workflows and work.
- (c) **Which of these features did you explore?** If you’re unsure of the list of features, you might learn more here: <https://oasis-lab.gitbook.io/roamresearch-discourse-graph-extension/> (*Options for each feature: “Didn’t know about this”, “Never used”, “Use on occasion”, “Core part of my workflow”*)
  - Node menu (for creating nodes)



- Discourse context
  - Query drawer (ad-hoc)
  - Discourse overlay
  - Grammar editor (for creating/modifying nodes and relation patterns)
  - Export
  - Playground
  - Node discourse attributes
  - Node index
  - Node templates
  - Query drawer (saved to pages)
- (d) **What were the main reasons you decided not to use the extension?** We are interested in understanding the main blockers to something like this extension being used by users like yourself, and what we might be able to do to remove those blockers. (*Free text*)

## B TECHNICAL DETAILS AND KEY DESIGN CHANGES

Figure 25 shows the grammar editor that shipped with the extension, which allowed users to add new discourse nodes (A), add new discourse relation patterns (B), and use a graphical layout to specify the underlying Datalog query in the Roam database that mapped to writing patterns that would enable the discourse graph extension to recognize the discourse relations (C) and also preview the writing patterns associated with the discourse relation pattern (D). As described in Section 6.2, user behavior led to significant changes that enabled more efficient editing of the grammar (Figure 26)

## C SUPPLEMENTARY DATA



Fig. 25. Snapshots from the grammar editor that shipped with the extension. Users could add new discourse nodes (A), add new discourse relation patterns (B), and use a graphical layout to specify the underlying Datalog query in the Roam database that mapped to writing patterns that would enable the discourse graph extension to recognize the discourse relations (C) and also preview the writing patterns associated with the discourse relation pattern (D)

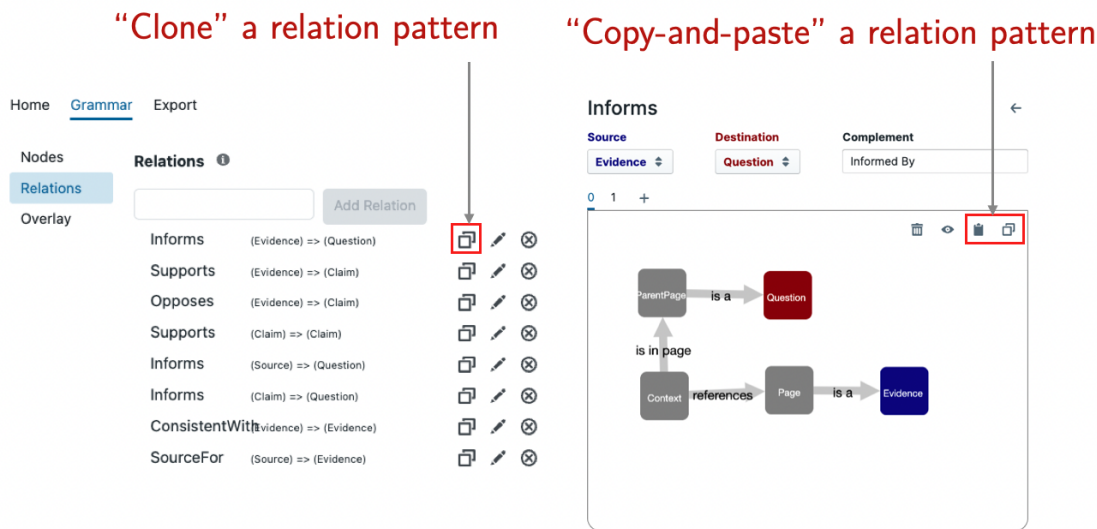


Fig. 26. Key design changes to the graph grammar editor — in response to user requests during the deploy period — to enable users to efficiently extend their discourse relation grammars by 1) “cloning” existing duplicate relation patterns, and 2) “copy-pasting” existing relation patterns.

	HCI lab	Cell biology lab (MA)	Medical imaging PhD student (LZ)	Criminology PhD student (GX)	Science Venture Capital Firm (DSV)	<i>Total</i>	<i>Average</i>
<b>QUESTION</b>	93	78	53	53		<b>370</b>	<b>69.3</b>
<b>CLAIM</b>	199	36	51	112		<b>597</b>	<b>99.5</b>
<b>EVIDENCE</b>	237	52	46	144	175	<b>891</b>	<b>130.8</b>
<b>SOURCE</b>	1751	225	352		2	<b>4081</b>	<b>582.5</b>
<b>EXPERIMENT</b>	14	5	2			<b>35</b>	<b>7.0</b>
<b>RESULT</b>	9	9	4			<b>31</b>	<b>7.3</b>
<b>HYPOTHESIS</b>		10	2			<b>12</b>	<b>6.0</b>
<b>METHOD</b>	3					<b>6</b>	<b>3.0</b>
<b>THEORY</b>	14			12		<b>40</b>	<b>13.0</b>
<b>PATTERN</b>	35			19		<b>89</b>	<b>27.0</b>
<b>ARTIFACT</b>	56					<b>112</b>	<b>56.0</b>
<b>ISSUE</b>	30	50	2			<b>112</b>	<b>27.3</b>
<b>CONCLUSION</b>		6	3			<b>9</b>	<b>4.5</b>
<b>COMPANY</b>					871	<b>871</b>	<b>871.0</b>
<b>CONSTRAINT</b>					267	<b>267</b>	<b>267.0</b>
<b>DEFINITION</b>				3		<b>3</b>	<b>3.0</b>
<b>HYPOTHESED_CONSTRAINTS</b>					606	<b>606</b>	<b>606.0</b>
<b>HYPOTHESED_SOLUTION</b>					1057	<b>1057</b>	<b>1057.0</b>
<b>SOLUTION</b>					264	<b>264</b>	<b>264.0</b>
<b>Total</b>	<b>2441</b>	<b>471</b>	<b>515</b>	<b>343</b>	<b>3242</b>	<b>9453</b>	<b>1402.4</b>

Fig. 27. Frequencies of each type of node created across 5 users who shared their discourse graph CSV export in the usage survey as of Fall 2022, after about 1 year of usage.