Open Access science needs Open Science Sensemaking (OSSm): open infrastructure for sharing scientific sensemaking data

Ronen Tamari[1,2] and Daniel A Friedman[2,3,4]

[1]Hebrew University of Jerusalem, Israel
[2]Common SenseMakers
[3]University of California, Davis, USA
[4]Active Inference Institute

April 26, 2023

Abstract

While open access publishing effectively broadens access to scientific research products, the problem of *making sense* of the volumes of new information is becoming increasingly acute. Traditional curation methods like peer-reviewed journals and recommendation services are failing to keep pace, resulting in unprecedented information overload and knowledge fragmentation. We contend that making sense of science requires open access to diverse sources of scientific sensemaking data, and that current frictions and failures of scientific sensemaking arise from deficiencies in reckoning with these kinds of data. Sensemaking data are the digital traces of sensemaking processes, in which individuals and groups organize and structure new information to improve subsequent decision-making and actions. Sensemaking data include explicit annotations (tags, votes, ratings, marginal comments) and commentaries made by researchers, as well implicit behavioral data generated through app usage (reference managers, website metrics, etc). Crucially, sensemaking data is currently scattered and siloed across a multitude of apps and formats, and also increasingly enclosed by publishers for profit. We provide an outline for Open Science Sensemaking (OSSm), an interoperable and decentralized annotation network. Such a system would enable researchers to record, own, and share their sensemaking data, thus contributing to the network while remaining resilient to platform capture. Shared annotation data will greatly benefit individual and collective sensemaking by enabling development of diverse content discovery services, from simple aggregation of reviews and ratings (e.g., "Goodreads for scientific research") to more advanced AI-augmented scientific intelligence systems.

1 Introduction

While recent developments in open access publishing and preprint servers like arXiv [24, 37] are valuable efforts that broaden access to scientific research, they do not address the challenge of making sense of all of the new information being published. Indeed, open access publishing may actually be amplifying this perennial challenge [12, 14] by dramatically increasing the amount of available published research without concomitant improvements in sensemaking practices, protocols and infrastructure [47, 58].

What do we mean by "making sense" of science? Some may question the use of this broad and seemingly vague term, preferring to refer specifically to more well-defined processes in science such as: literature discovery, technical evaluation, peer-review, meta-analysis, scientific communication, or scientometrics. We chose the sensemaking frame intentionally, to highlight what we see as a common root problem hindering progress on all of these important processes which are traditionally seen as disparate. The shared problem we want to draw attention to is a *lack of open access to sensemaking data generated by the scientific community.*[1]

[1]By scientific community, we refer not only to active scientists, but all people and entities that interact with scientific or scholarly research outputs more broadly. And while we focus on the scientific community and research processes in this article, many parallels exist in other professional and civic areas.

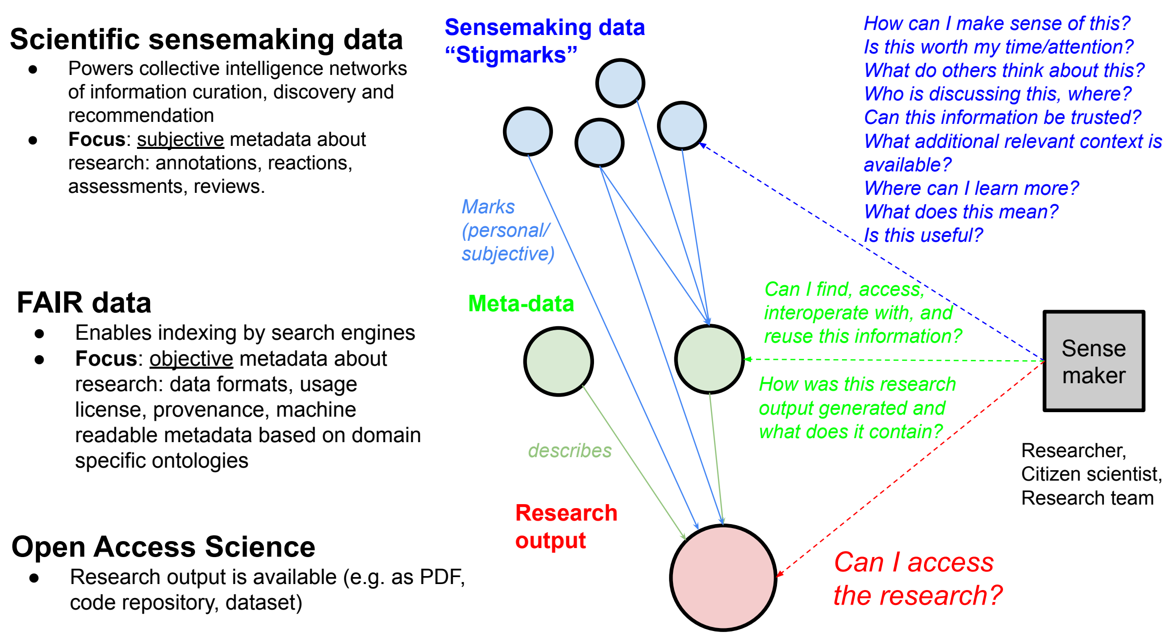

Figure 1: Overview of scientific sensemaking data (or "Stigmarks", blue circles on top) in relationship to FAIR data (here for meta-data, green circles in middle) and Open Access Science (research outputs, red circle at bottom). Edges reflect relationships among entities. A Sensemaker (gray square on the right side) represents active entities, who access or interact with these different types of informational entities.

Sensemaking refers to processes by which individuals and groups organize and structure new information to improve subsequent decision-making and actions [49, 52, 54, 68]. Sensemaking data are the digital traces of sensemaking processes, including explicit annotations made by researchers, as well implicit behavioral data generated by them through app usage. Sensemaking data represent reactions to existing content, rather than representing independent new content such as an article or code. Examples of explicit annotations include comments, ratings (numerical, votes, etc) and tags (topic, quality, reading priority etc). Example usage data includes download and viewership counts, click-through paths, and gaze and dwell-time data. The collection and use of modern sensemaking data is enabled by sensemaking infrastructure: tools for annotation and knowledge management, platforms, and other interfaces used by researchers for viewing, discovering, organizing, and sharing information. The importance of sensemaking data stems from their central role in algorithmic content curation and recommendation systems, where they function as valuable measurable proxies of relevance and attention [61].

Hence, sensemaking data and infrastructure are of central importance for the scientific community as well; as detailed in §2, they are instrumental for key meta-scientific challenges, including information overload, evaluation, training, scientometrics, and the support of large-scale distributed collaboration. However, the current infrastructure is failing to adequately address any of these challenges, and the sensemaking data required to make meaningful progress remains scattered across a multitude of platforms, apps and formats. As discussed in §3, some of these platforms enclose sensemaking data by design, and in other tools and platforms which are at least in principle more open, the lack of interoperable formats hinder efforts to effectively aggregate sensemaking data at scale.

Accordingly, we contend that in addition to the open science movement's efforts to ensure open access to scientific research, similar efforts must be directed at ensuring FAIR [46, 69] and open access to the sensemaking data that will allow us to make sense of that research (§4).

As an initial step towards open access sensemaking data, in this work we outline a proposal for Open Science Sensemaking, a decentralized annotation network enabling researchers to record, own and share their sensemaking data (§5). Our proposal focuses on a simple yet particularly important type of sensemaking process, Scientific Social Bookmarking (SSB); that of managing and sharing reading lists along with ratings and reviews.

The data shared through Open Science Sensemaking will support an ecosystem of diverse applications for collaborative creation, discovery and evaluation of knowledge. More broadly, infrastructure for sharing sensemaking data as a public good will play an instrumental role in re-integrating fragmented scientific knowledge, empowering researchers through technology, to think better, together.

2 Why care about scientific sensemaking data?

Sensemaking infrastructure and data are instrumental in addressing multiple key challenges for the scientific community, at both individual and collective levels:

Information overload. Individual scientists must keep up with unprecedented amounts of new research, and need better tools to help decide what papers to read next and where to focus their research attention [28]. Sensemaking data are essential for personalized algorithmic content recommendation, whether through generic social media networks [7] or bespoke scientific recommendation feeds [32].

Evaluation. Traditionally enacted in the form of journals and peer review, evaluation is vital for verification, curation and ranking of research outputs. As widely noted, the current peer review system is failing in all of these roles [21, 34, 35, 41]. Crowd-curated sensemaking signals (e.g., social media attention and usage data) are a promising part of the review and "pre-review" process, providing rapid and diverse types of feedback [15, 35]. Increasing access and quality of sensemaking data will improve open evaluation and help in steering attention towards work meriting more rigorous review.

Altmetrics. Altmetrics measure the broader impact of research and its public dissemination, to inform policy and funding decisions, as well as assessment of the contributions of individuals and institutions [51]. Where traditional impact metrics were based primarily on various citation analysis methods, newer approaches rely on signals curated from diverse online sources such as blog posts, social media mentions, and reference management tools like Zotero and Mendeley. However, standard altmetrics are critically hamstrung by the lack of open access to sensemaking data [19]. Open sensemaking data will improve the quality of altmetrics and diversity of measures that can be applied to track and assess research impact [22].

Coordinating big science. Scientific progress increasingly relies on large-scale efforts ("big science") [26, 31], and consequently individual scientists or small teams are less able to make meaningful progress [65]. Increased problem complexity necessitates distributed yet coordinated efforts across diverse groups and disciplines. Improving our infrastructure for open and shared annotation will play an important role in promoting cooperation and coordination, helping scientists find peers with similar interests, and quite literally helping to keep everyone on the same (digital) pages [15].

Modelling collective behavior of scientific communities. Similar to the importance of sensemaking data to the study of social systems more broadly [10, 33], scientific sensemaking data will be invaluable for meta-scientific research on attention allocation and information propagation in scientific communities [61].

To summarize, "lifting the hood" on key meta-scientific challenges reveals their reliance on sensemaking data. However, as described in the next section, at the same time we are seeing increasing enclosure and centralization of those data by commercial platforms.

3 Scientific sensemaking data, platformized

As observed by Chris Muellerleile in [47], the abundance of Open Access (OA) research brought with it new, yet underappreciated scientific attention economies: challenges of how to curate, filter, and make sense of the new research deluge. Muellerleile also notes that science publishers were quick to adapt to the new reality by transforming themselves into data aggregation and analytics platforms. Muellerleile concludes that "Open access to knowledge may be better than an environment where much academic knowledge is closed, but focusing too closely on the openness may be distracting us from the ways that capital is sneaking in the back door and enclosing the very tools we need to make sense of this new world." Those "very tools" being enclosed by publishers are, in large part, sensemaking data and the apps that generate them through user interactions. Similar to data enclosure patterns in social media platforms, sensemaking data are the "gold" of the attention economy, as they are essential for assessing reactions to content, as well as algorithmic curation, ranking and filtering [61]. Indeed, almost all scientific social bookmarking platforms (e.g., CiteULike and Connotea) and reference managers (e.g., Mendeley, Papers) were bought by large science publishers (e.g., Nature, Elsevier) for data tracking and analytics purposes [1, 2, 3, 4, 18, 50]. Twitter also recently cancelled a free data access service widely used by researchers [38].

More broadly, many works have made the case for a disconcerting *platformization* of scientific information infrastructure by commercial platforms [40, 50, 56]. Thus, despite ostensibly open access to publications, by amassing and enclosing data *about* publications (and especially sensemaking data), platforms have gained monopolistic power to determine the reach, value and impact of research. Similar to social media platforms where corporate self-interest takes precedence over pro-social incentives, commercial science platforms may be undermining scientific progress itself for the sake of profit [40, 41, 64].

4 Towards open, FAIR and stigmergic scientific sensemaking data

Given the importance of sensemaking data for realizing the potential of open access science, we contend that such data merit treatment and respect as scientific research outputs. In particular, as detailed below, we propose that scientific sensemaking data should be (1) included within the scope of the open science movement, (2) published under the FAIR principles (Findability, Accessibility, Interoperability, and Reuse), and (3) embedded in stigmergic annotation networks of content marking and discovery.

4.1 Open-access sensemaking data

Sensemaking data is different from traditional published scientific research in the sense that it is not public by default. In many cases, privacy considerations will limit the sensemaking data shared by researchers; they may not want to publicly share a particularly critical assessment, speculative claims, and so on (see §7.5 for further discussion). However, our focus is on settings in which researchers *do* wish to make their sensemaking data public, such as social bookmarking. For these cases, effective open infrastructure is currently lacking, leading researchers to this share data on closed commercial platforms such as Twitter or ResearchGate (§6.3).

4.2 FAIR sensemaking data

The FAIR Data Principles [46, 69] propose that research outputs should be Findable, Accessible, Interoperable, and Reusable. Here we consider the FAIR principles in the context of sensemaking data:

- **Findable.** Sensemaking data should be easily discoverable and indexed by search interfaces. We envision two primary kinds of queries: (1) querying a particular researcher/institution for their public "sensemaking activity feed"; what papers are they reading, research assessments, etc. (2) Querying content (such as a scientific paper) to retrieve associated sensemaking data such as reactions, annotations, reviews or ratings.
- **Accessible.** Sensemaking data should be retrievable using standardized protocols, while allowing for authentication procedures where necessary. Importantly, sensemaking data should be self-sovereign, with users controlling the accessibility of their data. As discussed above in relation to open access data (§4.1); to the extent that researchers wish to make their data public, they should be empowered to do so.
- **Interoperable.** Sensemaking data should use shared schemas for knowledge representation. Bookmarks are an example of a type of sensemaking data widely shared by researchers, but

scattered across many non-inter-operable formats, thus hindering integration efforts; for example, bookmarks may appear in an HTML reading list on a website, a public Zotero library, or a natural language post on social media.

- **Re-usable.** The creation context of the data and its provenance should be as rich as possible, to support accurate usage in downstream applications. Re-tweets are an example of sensemaking data provenance - researcher B's retweet of researcher A's post about paper P records the provenance of B's knowledge of P, and is valuable for applications such as content recommendation. Finally, data should be structured using expressive and machine actionable ontologies, to minimize ambiguity about what the data represents and support reliable use in downstream applications. For example, social media posts do not afford ontology-grounded representation: a researcher's sentiment towards a paper may be misinterpreted due to ambiguities of natural language.

4.3 Stigmergic sensemaking data

While FAIR data publishing is necessary, in the context of sensemaking, FAIR is not sufficient. Merely being findable in an ocean of content does not guarantee that information is distributed to the relevant parties or prioritized correctly with respect to other information. For these purposes, sensemaking data must be combined with personalizable algorithmic ranking and distribution systems. We adopt the concept of *stigmergic annotations* from [61] to refer to sensemaking data used to power large-scale collective intelligence networks through content discovery and algorithmic recommendation.[2] Importantly, stigmergic annotation networks incorporate feedback loops of content marking (annotation) and discovery which also provide social and epistemic participation incentives. In contrast to stigmergic annotations, most current annotation tools are not geared towards large networks, but are designed mainly for personal note-taking or small group use (§6.2). Twitter is an example of a stigmergic annotation network, demonstrating the power of social incentives along with simple annotations over other users (e.g., follow relations) and over content (likes, retweets, etc.) to drive large-scale personalized content recommendation. Notably, the recent open-sourcing of Twitter's recommendation algorithm demonstrates the central role of user social graphs and their content annotations (also called interaction data), without need for analysis over the actual content of posts [7]. Notwithstanding platforms' impressive potential, their centralized and closed nature motivates us to propose an open and decentralized scientific sensemaking network, as described next.

5 Proposal: Open Science Sensemaking (OSSm)

Our proposal, Open Science Sensemaking, builds on the recent Open Source Attention initiative [61], which proposed a decentralized stigmergic annotation network to share sensemaking data as a public good, as an alternative to its capture by commercial social media platforms. Such a network could be further specified using systems engineering tools such as the Active Entity Ontology for Science (AEOS) [22]. In this case, sensemaking data would be reflected by various kinds of informational artefacts, and active entities would correspond to different scientific sensemakers (people, teams, organizations, etc).

5.1 Open Science Sensemaking Graph

The core envisioned network infrastructure is the Open Science Sensemaking Graph (OSSG), a shared public knowledge graph representing sensemaking processes over scientific research. We plan to focus initially on the setting of social bookmarking, with later evolution of the network incorporating more diverse sensemaking processes (§5.4). As shown in Figure 2, OSSG enriches the traditional academic citation graph with a sensemaking layer: beyond authors and papers, sensemaking data are also represented. More concretely, the graph will include 3 main types of nodes:

1. **Actors (Sensemakers).** Human individuals (scientists, general public) or collectives (organizations, research teams), as well as machine agents. Actors jointly construct the Open Science Sensemaking Graph by creating **stigmergic markers** (see below). Following [61], we also refer

[2] Stigmergy refers to mechanisms of indirect communication between agents mediated by marks they leave in their environment (e.g., ant pheromone trails). For details, see [45, 61].

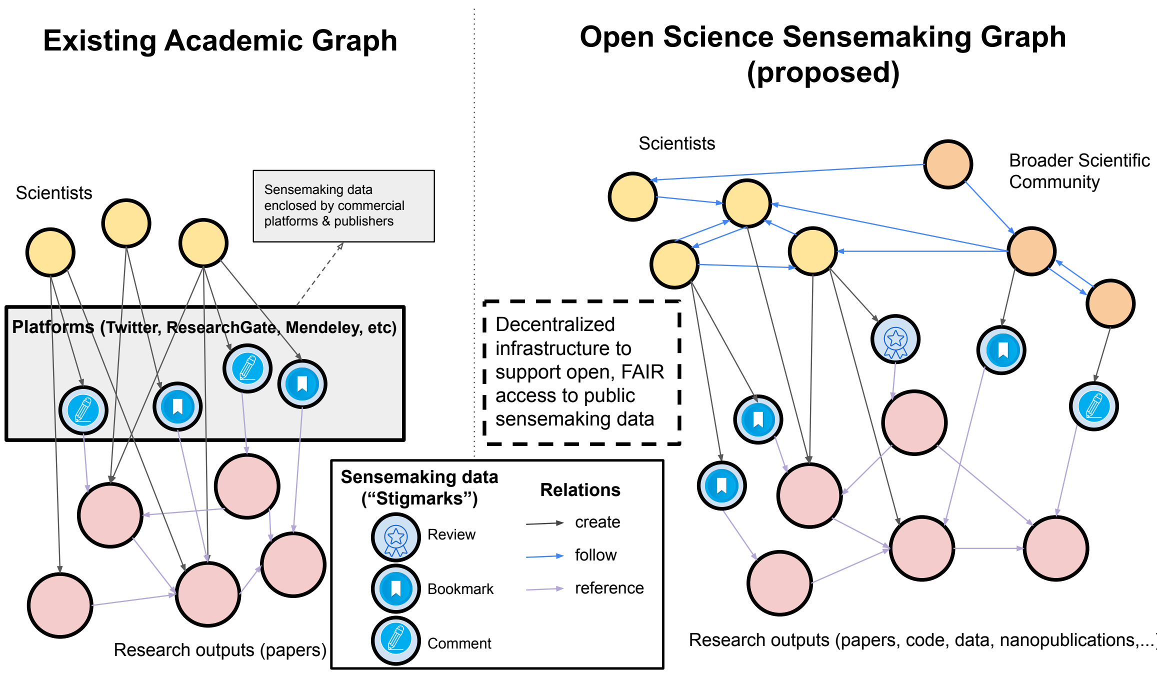

Figure 2: Proposed Open Science Sensemaking Graph (right) shown in comparison to standard academic graph representation (left). Our proposal includes an open access layer of public sensemaking data over the existing content, in contrast to the current situation in which only the academic graph is open, and sensemaking data is enclosed and fragmented across multiple platforms.

to actors as SenseMakers, to highlight their active role in collective sensemaking, in contrast to social network "users".

2. **Content (Scientific research).** Any Web-addressable content, such as websites or social media posts. For simplicity, any web resource with a URL can be treated as a content node.[3] To sidestep issues around copyrights, the content itself is not represented, but only the address of the content (e.g., the URL).

3. **Stigmergic markers (Sensemaking data).** Stigmergic markers (**StigMarks** or **marks** for short) include signals or annotations over content or other actors. We refer to them as stigmergic to differentiate them from personal annotations: stigmergic markers are those made primarily for social sensemaking purposes, "trail marks" or "digital pheromone trails" to help others navigate information.

An example use case of the above, is shown in Figure 3.

5.1.1 Types of StigMarks

Broadly, we distinguish between marks over actors, and marks over content.

Marks over content. For the bookmarking setting, the main mark actors can create is a simple `bookmark`, represented by the following fields (some may be optional):

- Target URL
- Comments (short text)
- Tags for categorizing content by topics
- Reading status (`to-read`, `reading`, `finished-reading`)

[3]Future versions should include more rigorous knowledge graph maintenance including entity disambiguation, version tracking, and linking

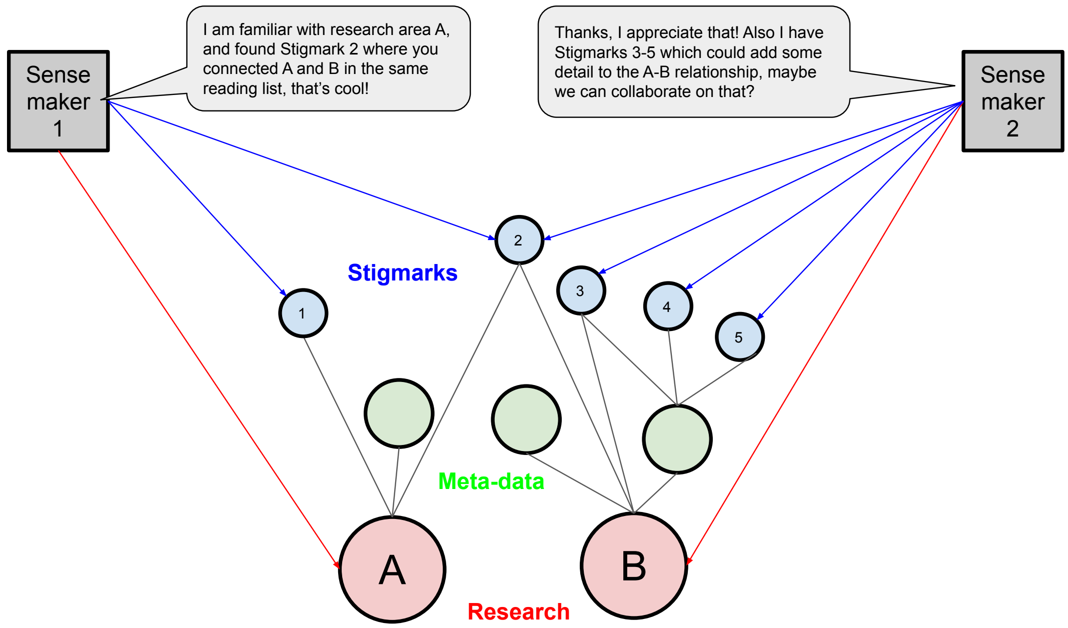

Figure 3: Example use case of Open Science Sensemaking (OSSm). Two sensemakers connect and engage in a conversation supported by shared access to relevant sensemaking data (Stigmarks).

- Rating. For a start, this could be a simple scalar scale. Future developments could include the multi-dimensoinal rating scheme proposed by [35].

Marks over actors. Initially, the main mark over actors is the actor-actor `follow` relation.[4] Similar to its counterparts on social media networks, `follow(A,B)` indicates user A is interested in seeing content created or recommended by B. The social graph structure created by `follow` relations is an important signal to recommendation algorithms as well as a way to induce trust networks and increase resilience to spamming and other attacks [30]. Future versions can include additional types of actor-actor relations, such as expressions of endorsement or trust [29].

5.2 Protocol-driven decentralized data-sharing

The network will be built using a protocol based approach, using standardized data schemas to enable interoperability across apps. Broadly, the network architecture is inspired by the Murmurations protocol for decentralized, interoperable data-sharing (for further details see §5.3, [6]). The network will consist of 4 main modules:

1. Data storage. Data storage will be decentralized to prevent the data being acquired and enclosed: ideally, each member in the network will host and manage their own data, using Solid data pods [55], IPFS [11] or other distributed storage technologies. Federated/hosted approaches may also be employed as a quick start option or to serve users that prefer to outsource data storage.

2. Indexing and querying services. For data to be useful for collective sensemaking applications, it must be indexed and queryable by various apps. Members will register their bookmarks with an indexing service to allow their data to be discovered by the network.

3. Content discovery services. A variety of apps leveraging the shared bookmarks data and indexing/querying services to provide search and recommendation over content, people and organizations.

[4]For simplicity, pure relations like `follow` can also be expressed by edges and do not need to be represented by a node.

4. User-facing apps. Apps allowing users to easily create and organize marks, and control data visibility.

Progressive decentralization. Open Science Sensemaking will follow a progressive decentralization strategy [43]. While the ultimate goal is self-sovereign decentralized storage by network members, starting from centralized storage and progressively decentralizing can help strike a balance between the practicalities of software development and the ideals of decentralization.

5.3 Initial implementation sketch

Our proposed initial implementation will focus on a minimum viable version of the network. A simple user-facing bookmarking app will be built as a reference implementation, providing an interface to enter input data as well as handling network registration and storage configuration. Integrations with popular existing knowledge management tools (e.g., Obsidian, Notion) and reference managers (e.g., PaperPile, Zotero) will also be considered. A minimal yet useful content discovery service could be a simple numerical review and rating aggregator for research papers (similar to the data aggregation provided by Goodreads). This kind of information can be embedded directly in papers' landing pages on preprint/journal servers, similarly to other information tabs provided by external data sources (for example, arXiv provides various 3rd party "related papers" data). A simple initial recommender algorithm could be implemented using collaborative filtering and social graph data [35].

5.4 Beyond bookmarks

In future stages we also envision including additional annotation types to support additional sensemaking processes, such as span annotation (highlighting text) and semantic linking between documents or parts of documents (for example, that document A supports a claim in document B). Discourse graphs for knowledge synthesis are a promising sensemaking process supported by semantic linking in and across documents [17]. Finally, integrations with nanopublications (§6.1) can introduce a wide range of ontology-based semantic knowledge representations.

6 Related work

6.1 Nanopublications

Nanopublications are being developed to communicate the smallest units of publishable information in an expressive and machine readable format, using FAIR data principles [25, 36]. The Nanopublications project shares much in common with Open Science Sensemaking, as both are aimed at developing open and collaborative infrastructure to help researchers find, curate and connect scientific knowledge. Stigmarks are conceptually similar to nanopublications, and indeed there is an overlap with some existing nanopublication templates, such as an assertion of reading a paper[5], or a review of a scientific paper.[6] While sharing similar broader goals and methods, nanopublications are focused more on formal knowledge and discourse representation, and Open Science Sensemaking is more focused on the challenges of information overload, stigmergic annotation and data-driven collective knowledge curation. Our initiatives are naturally synergistic: as noted by [23], nanopublications provide an expressive and robust data representation substrate, but additionally require a diverse ecosystem of content discovery and authoring tools to more fully realize their potential. Our proposal additionally focuses on the social and epistemic incentives for sharing sensemaking data, which we believe will be essential for driving large scale use (4.3).

6.2 Annotation interfaces

There are a wide array of projects focused on annotation tools, including open source platforms like hypothes.is, as well as tools allowing users to self-host their annotations, such as dokieli [16]. While

[5] http://purl.org/np/RAxPdvy5RN-jyPOMcBNEsUEn2CPBtAa3W0Ct3tbID4PiM
[6] http://purl.org/np/RA1sViVmXf-W2aZW4Qk74KTaiD9gpLBPe2LhMsinHKKz8

some of these tools support a degree of FAIR publishing, they are designed for individuals and small-groups rather than for large scale stigmergic sensemaking (4.3). For example, they do not support user interactions with annotations (votes, re-tweets, etc) or other easily aggregatable annotations like ratings. Unfold Research [8] is another related app, that does allow directly annotating pages with relevant aggregatable data such as up/down-votes, but does not currently support FAIR data publishing. Finally, these apps also do not integrate the social networks necessary for more personalized content recommendation and social incentives.

6.3 Social media platforms

Scientists increasingly use social media platforms like Twitter and Mastodon extensively for sharing comments, reviews and recommendations [42]. While demonstrating the great value of digital social networks for catalyzing scientific research, these platforms are also limited by lack of support for FAIR data. Posts are limited to natural language which harms machine readability. Closed commercial social media platforms like Twitter operate opaque recommendation algorithms and are also at risk of data enclosure. Finally, many researchers prefer to avoid the exposure and distractions of social media platforms.

6.4 AI-based search and recommendation

Artificial Intelligence (AI) research has made impressive progress, particularly large language models (LLMs) for open domain dialog [48, 67]. Ellicit, Semantic Scholar, and many more companies are providing AI and LLM-driven services for content recommendation, question answering and summarization over scientific research [32, 53, 62]. While various algorithmic dialog, search and recommendation systems will be indispensable parts of any solution [28], they are still only partial solutions [57]: collective sensemaking is ultimately a social human process [13, 59]. For example, while recommendation systems may suggest interesting papers, researchers naturally want to know what papers their human peers and role-models are reading [39]. Furthermore, as demonstrated by researchers' use of social media platforms, trust and social incentives drive researchers to share information, discuss papers and seek engagement with their peers. Finally, open access to sensemaking data is also a concern for AI systems due to their reliance on large scale data sources. Recommendation systems in particular require high quality human sensemaking data to produce good recommendations (§4.3).

6.5 Decentralized ontology-agnostic data sharing protocols

Another related effort is the Murmurations protocol [6]. Murmurations is designed to allow users to manage and share their own data in an ontology agnostic way, meaning that any data schema can be supported, resulting in a high degree of interoperability. For effective sharing and content discovery, Murmurations also defines indexing and aggregation. Murmurations has thus far been applied to geo-spatial data and directories of communities - our proposal would expand upon it to support data related to content bookmarking and social graph information.

6.6 Decentralized Science

Decentralized science ("DeSci") is a broad term encompassing a range of projects aiming to build public infrastructure for funding, creating, reviewing, crediting, storing, and disseminating scientific knowledge fairly and equitably using decentralized technologies such as distributed ledgers (blockchain) [5, 9, 22, 63, 66]. In common with our proposal, DeSci highlights open access, FAIR and decentralized publishing of scientific knowledge. To our knowledge though, in DeSci, similar to the situation in open access science, little work has focused on open sensemaking infrastructure: tools and protocols for collaborative content curation, annotation, discovery and recommendation.

7 Discussion

This section briefly touches upon key challenges and open questions related to our proposal; we leave a more thorough exploration to future work.

7.1 Starting and Scaling

As common in networked information systems, lack of data in initial stages discourages wider user adoption, limiting the growth of the system. Bootstrapping from existing data sources is a potential way to address this challenge. Possible bootstrapping approaches include providing a variety of integrations with existing bookmarking apps, and also automated text mining methods to extract relevant data from existing social networks such as Twitter (while accounting also for user consent).

7.2 Adoption

Another related challenge is community uptake; e.g., convincing researchers to adopt new tools and practices regarding data sharing. An important way to address this challenge is through "meeting users where they are", or providing integrations with existing popular pre-print servers, note taking and bookmarking apps, in order to reduce friction. Also important is providing clear utility for participating in the network, for example in the form of new content discovery options and social incentives [60, 70].

7.3 Discoverability

Beyond the challenge of adopting new tools, for the collected annotations to be useful, they must be easily discoverable by researchers. One approach could be embedding bookmarking data on the landing pages of research papers in high-profile publications and popular preprint servers like arXiv, similar to how social media mentions are currently displayed (and see also initial experiments with nanopublications[7]).

7.4 Information security

Enabling permissionless contribution to public data pools brings the risk of spam or misinformation by malicious actors. Adding blockchain based credentialing or social graphs can help counter these threats, by filtering out data from untrusted sources. More broadly, these threats to epistemic integrity, and associated proactive/responsive practices, can be considered from the perspective of cognitive security [20].

7.5 Privacy

To emphasize, our proposal is focused on the *public* FAIR sharing of sensemaking data which lacks viable alternatives to commercial platforms, as opposed to private or group sharing which are already well-supported (§6.2). As noted by [70], most work has focused on protecting user data from involuntary tracking, rather than negotiating the cost-benefit tradeoffs involved in voluntary public sharing of sensemaking data; balancing privacy risks on one hand, with benefits such as gains in information and social capital on the other. Important considerations to account for in this setting include providing users with intuitive control over data visibility settings, and opt-in publishing models requiring explicit user consent to make data public. In future iterations we plan to add the option for group-level visibility settings. Anonymous and privacy preserving data sharing options should also be explored.

7.6 Business model

As discussed in §3, technical and financial dimensions of platforms are entangled; platforms rely on data enclosure as a source of revenue. Accordingly, beyond technological solutions, another key challenge is implementing a business model that protects against data enclosure. Various possible alternatives to current platform models could be explored, including:

1. Subscription/freemium models, where some services are freely available but more advanced apps require payment.

[7] https://mstdn.social/@RIOjournal/110185219008689466

2. App marketplace monetization. The open data model encourages a diverse app market providing a variety of services, such as search, content recommendation and social matching (pairing readers with similar interests, etc). App markets are another potential source of revenue- when the user purchases a 3rd-party app, a portion of the proceeds would be charged by the company. Notably, subscriptions and an app marketplace are the two main sources of revenue for the popular code sharing platform Github.

3. Data marketplace. Another potentially promising model is a data co-operative [27], where users contribute to a shared data pool and are also paid for access to their data. Here too, the organization could charge a fee in facilitating the matching between data owners and buyers.

4. Grants. There is increasing awareness around the limitations of existing knowledge infrastructures, and many DeSci as well as traditional organizations are offering grant-based funding for related initiatives.

5. Focused Research Organization (FRO). FROs [44] are a new kind of special-purpose organization created to solve scientific or technological challenges that are not efficiently addressed by the existing organizational structures of academia, industry, or government. FROs are mission oriented and shielded from both academic and for-profit incentives, making them a promising potential model for large-scale metascience projects such as Open Science Sensemaking.[8]. By analogy to FROs, if open access to scientific sensemaking data were in place with distributed protocols, "Focused Sensemaking Organizations" (FSO) would be relevant in future scientific settings.

References

[1] Connotea to discontinue service. https://blogs.nature.com/ofschemesandmemes/2013/01/24/connotea-to-discontinue-service, 2013. Accessed: 2023-04-03.

[2] Atypon, elsevier and digital science collaborate to deliver usage reporting. https://www.atypon.com/news/atypon-elsevier-and-digital-science-collaborate-to-deliver-usage-reporting/, 2018. Accessed: 2023-04-25.

[3] Citeulike is closing down. https://web.archive.org/web/20190310145602/http://www.citeulike.org/news, 2019. Accessed: 2023-04-03.

[4] Welcome to hotel elsevier: you can check-out any time you like … not. https://eiko-fried.com/welcome-to-hotel-elsevier-you-can-check-out-any-time-you-like-not/, 2022. Accessed: 2023-04-04.

[5] Desci labs - tools to grow open science. https://www.desci.com/about, 2023. Accessed: 2023-04-19.

[6] Murmurations - making movements visible. https://murmurations.network/, 2023. Accessed: 2023-04-10.

[7] Twitter showed us its algorithm. what does it tell us? https://knightcolumbia.org/blog/twitter-showed-us-its-algorithm-what-does-it-tell-us, 2023. Accessed: 2023-04-17.

[8] Unfold research. hhttps://unfoldresearch.com/, 2023. Accessed: 2023-04-19.

[9] A guide to desci, the latest web3 movement. https://a16zcrypto.com/content/article/what-is-decentralized-science-aka-desci/, 2023. Accessed: 2023-04-19.

[10] Joseph B. Bak-Coleman, Mark Alfano, Wolfram Barfuss, Carl T. Bergstrom, Miguel A. Centeno, Iain D. Couzin, Jonathan F. Donges, Mirta Galesic, Andrew S. Gersick, Jennifer Jacquet, Albert B. Kao, Rachel E. Moran, Pawel Romanczuk, Daniel I. Rubenstein, Kaia J. Tombak, Jay J. Van Bavel, and Elke U. Weber. Stewardship of global collective behavior. *Proceedings of the National Academy of Sciences*, 118(27):e2025764118, 2021. doi: 10.1073/pnas.2025764118. URL https://www.pnas.org/doi/abs/10.1073/pnas.2025764118.

[8] https://astera.org/metascience/

[11] Juan Benet. Ipfs - content addressed, versioned, p2p file system, 2014.

[12] Ann M Blair. *Too much to know: Managing scholarly information before the modern age*. Yale University Press, 2010.

[13] Andrew D. Brown, Patrick Stacey, and Joe Nandhakumar. Making sense of sensemaking narratives. *Human Relations*, 61(8):1035–1062, 2008. doi: 10.1177/0018726708094858. URL https://doi.org/10.1177/0018726708094858. _eprint: https://doi.org/10.1177/0018726708094858.

[14] Vannevar Bush. As we may think. *The atlantic monthly*, 176(1):101–108, 1945.

[15] Brett T. Buttliere. Using science and psychology to improve the dissemination and evaluation of scientific work. *Frontiers in Computational Neuroscience*, 8, 2014. ISSN 1662-5188. URL https://www.frontiersin.org/articles/10.3389/fncom.2014.00082.

[16] Sarven Capadisli. *Linked research on the decentralised Web*. PhD thesis, 2020. URL https://csarven.ca/linked-research-decentralised-web.

[17] Joel Chan, Wayne Lutters, Jodi Schneider, Karola Kirsanow, Silvia Bessa, and Jonny L. Saunders. Growing new scholarly communication infrastructures for sharing, reusing, and synthesizing knowledge. In *Companion Publication of the 2022 Conference on Computer Supported Cooperative Work and Social Computing*, CSCW'22 Companion, page 278–281, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450391900. doi: 10.1145/3500868.3559398. URL https://doi.org/10.1145/3500868.3559398.

[18] Leslie Chan and Pierre Mounier, editors. *Connecting the Knowledge Commons — From Projects to Sustainable Infrastructure : The 22nd International Conference on Electronic Publishing – Revised Selected Papers*. Laboratoire d'idées. OpenEdition Press, Marseille, June 2019. ISBN 979-10-365-3802-5. doi: 10.4000/books.oep.8999. URL http://books.openedition.org/oep/8999. Code: Connecting the Knowledge Commons — From Projects to Sustainable Infrastructure : The 22nd International Conference on Electronic Publishing – Revised Selected Papers Publication Title: Connecting the Knowledge Commons — From Projects to Sustainable Infrastructure : The 22nd International Conference on Electronic Publishing – Revised Selected Papers Reporter: Connecting the Knowledge Commons — From Projects to Sustainable Infrastructure : The 22nd International Conference on Electronic Publishing – Revised Selected Papers Series Title: Laboratoire d'idées.

[19] Kathy Christian, Euan Adie, Gemma Derrick, Fereshteh Didegah, Paul Groth, Cameron Neylon, Jason Priem, Shenmeng Xu, Zohreh Zahedi, Y-L Theng, et al. The state of altmetrics: a tenth anniversary celebration. *The state of altmetrics: a tenth anniversary celebration*, 2020.

[20] Cordes RJ. David, S. and DA. Friedman. *Structuring the Information Commons: Open Standards and Cognitive Security*. Cognitive Security Education Forum (COGSEC), 2022. ISBN 978-1-7364269-3-7. URL https://www.cogsec.org/cat-22.

[21] Michael B Eisen, Anna Akhmanova, Timothy E Behrens, Jörn Diedrichsen, Diane M Harper, Mihaela D Iordanova, Detlef Weigel, and Mone Zaidi. Scientific Publishing: Peer review without gatekeeping. *eLife*, 11:e83889, October 2022. ISSN 2050-084X. doi: 10.7554/eLife.83889. URL https://doi.org/10.7554/eLife.83889. Publisher: eLife Sciences Publications, Ltd.

[22] Daniel Friedman, Shaun Applegate-Swanson, Arhan Choudhury, RJ Cordes, Shady El Damaty, Avel Guénin—Carlut, V. Bleu Knight, Ivan Metelkin, Siddhant Shrivastava, Amit Kumar Singh, Jakub Smékal, Tuttle. Caleb, and Alexander Vyatkin. An Active Inference Ontology for Decentralized Science: from Situated Sensemaking to the Epistemic Commons, March 2022. URL https://doi.org/10.5281/zenodo.6320575.

[23] Fabio Giachelle, Dennis Dosso, and Gianmaria Silvello. Search, access, and explore life science nanopublications on the Web. *PeerJ Computer Science*, 7:e335, February 2021. ISSN 2376-5992. doi: 10.7717/peerj-cs.335. URL https://peerj.com/articles/cs-335. Publisher: PeerJ Inc.

[24] Paul Ginsparg. ArXiv at 20. *Nature*, 476(7359):145–147, August 2011. ISSN 1476-4687. doi: 10.1038/476145a. URL https://www.nature.com/articles/476145a. Number: 7359 Publisher: Nature Publishing Group.

[25] Paul Groth, Andrew Gibson, and Jan Velterop. The anatomy of a nanopublication. *Information Services & Use*, 30(1-2):51–56, January 2010. ISSN 0167-5265. doi: 10.3233/ISU-2010-0613. URL https://content.iospress.com/articles/information-services-and-use/isu613. Publisher: IOS Press.

[26] Olof Hallonsten. *Big science transformed*. Springer, 2016.

[27] Thomas Hardjono and Alex Pentland. Data cooperatives: Towards a foundation for decentralized personal data management, 2019.

[28] Tom Hope, Doug Downey, Oren Etzioni, Daniel S. Weld, and Eric Horvitz. A computational inflection for scientific discovery, 2022.

[29] Farnaz Jahanbakhsh, Amy X. Zhang, and David R. Karger. Leveraging structured trusted-peer assessments to combat misinformation. *Proc. ACM Hum.-Comput. Interact.*, 6(CSCW2), nov 2022. doi: 10.1145/3555637. URL https://doi.org/10.1145/3555637.

[30] Wenjun Jiang, Guojun Wang, Md Zakirul Alam Bhuiyan, and Jie Wu. Understanding graph-based trust evaluation in online social networks: Methodologies and challenges. *ACM Comput. Surv.*, 49(1), may 2016. ISSN 0360-0300. doi: 10.1145/2906151. URL https://doi.org/10.1145/2906151.

[31] Benjamin F. Jones. As Science Evolves, How Can Science Policy? *Innovation Policy and the Economy*, 11:103–131, January 2011. ISSN 1531-3468. doi: 10.1086/655820. URL https://www.journals.uchicago.edu/doi/full/10.1086/655820. Publisher: The University of Chicago Press.

[32] Hyeonsu B Kang, Rafal Kocielnik, Andrew Head, Jiangjiang Yang, Matt Latzke, Aniket Kittur, Daniel S Weld, Doug Downey, and Jonathan Bragg. From who you know to what you read: Augmenting scientific recommendations with implicit social networks. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450391573. doi: 10.1145/3491102.3517470. URL https://doi.org/10.1145/3491102.3517470.

[33] Anne Kohlbrenner, Ben Kaiser, Kartikeya Kandula Weiss, Jonathan Mayer, Ted Han, Robert Helmer, et al. Rally and webscience: A platform and toolkit for browser-based research on technology and society problems. *arXiv preprint arXiv:2211.02274*, 2022.

[34] Michail Kovanis, Raphaël Porcher, Philippe Ravaud, and Ludovic Trinquart. The Global Burden of Journal Peer Review in the Biomedical Literature: Strong Imbalance in the Collective Enterprise. *PLOS ONE*, 11(11):1–14, November 2016. doi: 10.1371/journal.pone.0166387. URL https://doi.org/10.1371/journal.pone.0166387. Publisher: Public Library of Science.

[35] Nikolaus Kriegeskorte. Open Evaluation: A Vision for Entirely Transparent Post-Publication Peer Review and Rating for Science. *Frontiers in Computational Neuroscience*, 6, 2012. ISSN 1662-5188. doi: 10.3389/fncom.2012.00079. URL https://www.frontiersin.org/articles/10.3389/fncom.2012.00079.

[36] Tobias Kuhn, Paolo Emilio Barbano, Mate Levente Nagy, and Michael Krauthammer. Broadening the Scope of Nanopublications. In Philipp Cimiano, Oscar Corcho, Valentina Presutti, Laura Hollink, and Sebastian Rudolph, editors, *The Semantic Web: Semantics and Big Data*, Lecture Notes in Computer Science, pages 487–501, Berlin, Heidelberg, 2013. Springer. ISBN 978-3-642-38288-8. doi: 10.1007/978-3-642-38288-8_33.

[37] Mikael Laakso, Patrik Welling, Helena Bukvova, Linus Nyman, Bo-Christer Björk, and Turid Hedlund. The Development of Open Access Journal Publishing from 1993 to 2009. *PLOS ONE*, 6(6):1–10, June 2011. doi: 10.1371/journal.pone.0020961. URL https://doi.org/10.1371/journal.pone.0020961. Publisher: Public Library of Science.

[38] Heidi Ledford. Researchers scramble as Twitter plans to end free data access. *Nature*, 614(7949): 602–603, February 2023. doi: 10.1038/d41586-023-00460-z. URL https://www.nature.com/articles/d41586-023-00460-z. Bandiera_abtest: a Cg_type: News Number: 7949 Publisher: Nature Publishing Group Subject_term: Scientific community, Sociology, Technology.

[39] Yuchen Liu, Dmitry Chechik, and Junghoo Cho. Power of Human Curation in Recommendation System. In *Proceedings of the 25th International Conference Companion on World Wide Web*, WWW '16 Companion, pages 79–80, Republic and Canton of Geneva, CHE, April 2016. International World Wide Web Conferences Steering Committee. ISBN 978-1-4503-4144-8. doi: 10.1145/2872518.2889350. URL https://dl.acm.org/doi/10.1145/2872518.2889350.

[40] Lai Ma. Information, platformized. *Journal of the Association for Information Science and Technology*, 74(2):273–282, 2023. doi: https://doi.org/10.1002/asi.24713. URL https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/asi.24713. _eprint: https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/asi.24713.

[41] Stuart Macdonald. The gaming of citation and authorship in academic journals: a warning from medicine. *Social Science Information*, page 05390184221142218, February 2023. ISSN 0539-0184. doi: 10.1177/05390184221142218. URL https://doi.org/10.1177/05390184221142218. Publisher: SAGE Publications Ltd.

[42] Merja Mahrt, Katrin Weller, and Isabella Peters. Twitter in scholarly communication. *Twitter and society*, 89:399–410, 2014.

[43] Morshed Mannan and Nathan Schneider. Exit to community: Strategies for multi-stakeholder ownership in the platform economy. *Geo. L. Tech. Rev.*, 5:1, 2021.

[44] Adam Marblestone, Anastasia Gamick, Tom Kalil, Cheryl Martin, Milan Cvitkovic, and Samuel G. Rodriques. Unblock research bottlenecks with non-profit start-ups. *Nature*, 601(7892):188–190, January 2022. doi: 10.1038/d41586-022-00018-5. URL https://www.nature.com/articles/d41586-022-00018-5. Bandiera_abtest: a Cg_type: Comment Number: 7892 Publisher: Nature Publishing Group Subject_term: Research management, Institutions.

[45] Leslie Marsh and Christian Onof. Stigmergic epistemology, stigmergic cognition. *Cognitive Systems Research*, 9(1-2):136–149, 2008. ISSN 13890417. doi: 10.1016/j.cogsys.2007.06.009.

[46] Barend Mons, Cameron Neylon, Jan Velterop, Michel Dumontier, Luiz Olavo Bonino da Silva Santos, and Mark D. Wilkinson. Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. *Information Services & Use*, 37(1):49–56, January 2017. ISSN 0167-5265. doi: 10.3233/ISU-170824. URL https://content.iospress.com/articles/information-services-and-use/isu824. Publisher: IOS Press.

[47] Chris Muellerleile. Open access panacea: Scarcity, abundance, and enclosure in the new economy of academic knowledge production 1. In *The Routledge handbook of the political economy of science*, pages 132–143. Routledge, 2017.

[48] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Gray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. Training language models to follow instructions with human feedback. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho, editors, *Advances in Neural Information Processing Systems*, 2022. URL https://openreview.net/forum?id=TG8KACxEON.

[49] Peter Pirolli and Stuart Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. pages 2–4, 2005. URL https://analysis.mitre.org/proceedings/Final_Papers_Files/206_Camera_Ready_Paper.pdf.

[50] Jean-Christophe Plantin, Carl Lagoze, and Paul N Edwards. Re-integrating scholarly infrastructure: The ambiguous role of data sharing platforms. *Big Data & Society*, 5(1):2053951718756683, January 2018. ISSN 2053-9517. doi: 10.1177/2053951718756683. URL https://doi.org/10.1177/2053951718756683. Publisher: SAGE Publications Ltd.

[51] Jason Priem, Dario Taraborelli, Paul Groth, and Cameron Neylon. Altmetrics: A manifesto. 2011.

[52] Gonzalo Ramos, Napol Rachatasumrit, Jina Suh, Rachel Ng, and Christopher Meek. ForSense: Accelerating Online Research Through Sensemaking Integration and Machine Research Support. *ACM Transactions on Interactive Intelligent Systems*, 12(4):30:1–30:23, November 2022. ISSN 2160-6455. doi: 10.1145/3532853. URL https://dl.acm.org/doi/10.1145/3532853.

[53] Justin Reppert, Ben Rachbach, Charlie George, Luke Stebbing, Jungwon Byun, Maggie Appleton, and Andreas Stuhlmüller. Iterated decomposition: Improving science qa by supervising reasoning processes, 2023.

[54] Daniel M. Russell, Mark J. Stefik, Peter Pirolli, and Stuart K. Card. The cost structure of sensemaking. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems*, CHI '93, pages 269–276, New York, NY, USA, May 1993. Association for Computing Machinery. ISBN 978-0-89791-575-5. doi: 10.1145/169059.169209. URL https://dl.acm.org/doi/10.1145/169059.169209.

[55] Andrei Vlad Sambra, Essam Mansour, Sandro Hawke, Maged Zereba, Nicola Greco, Abdurrahman Ghanem, Dmitri Zagidulin, Ashraf Aboulnaga, and Tim Berners-Lee. Solid: a platform for decentralized social applications based on linked data.

[56] Jonny L. Saunders. Decentralized infrastructure for (neuro)science. 2022. doi: 10.48550/ARXIV.2209.07493. URL https://arxiv.org/abs/2209.07493.

[57] Chirag Shah and Emily M. Bender. Situating search. In *ACM SIGIR Conference on Human Information Interaction and Retrieval*, CHIIR '22, page 221–232, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450391863. doi: 10.1145/3498366.3505816. URL https://doi.org/10.1145/3498366.3505816.

[58] Steinn Sigurdsson. The future of arXiv and knowledge discovery in open science. In *Proceedings of the First Workshop on Scholarly Document Processing*, pages 7–9, Online, November 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.sdp-1.2. URL https://aclanthology.org/2020.sdp-1.2.

[59] David Snowden. The social ecology of knowledge management. In *Knowledge horizons*, pages 237–265. Routledge, 2012.

[60] Sue Yeon Syn and Sanghee Oh. Why do social network site users share information on Facebook and Twitter? *Journal of Information Science*, 41(5):553–569, 2015. doi: 10.1177/0165551515585717. URL https://doi.org/10.1177/0165551515585717. _eprint: https://doi.org/10.1177/0165551515585717.

[61] Ronen Tamari, Daniel Friedman, William Fischer, Lauren Hebert, and Dafna Shahaf. From users to (sense)makers: On the pivotal role of stigmergic social annotation in the quest for collective sensemaking. In *Proceedings of the 33rd ACM Conference on Hypertext and Social Media*, HT '22, page 236–239, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450392334. doi: 10.1145/3511095.3536361. URL https://doi.org/10.1145/3511095.3536361.

[62] Ross Taylor, Marcin Kardas, Guillem Cucurull, Thomas Scialom, Anthony Hartshorn, Elvis Saravia, Andrew Poulton, Viktor Kerkez, and Robert Stojnic. Galactica: A large language model for science, 2022.

[63] Ámbar Tenorio-Fornés, Elena Pérez Tirador, Antonio A. Sánchez-Ruiz, and Samer Hassan. Decentralizing science: Towards an interoperable open peer review ecosystem using blockchain. *Information Processing & Management*, 58(6):102724, 2021. ISSN 0306-4573. doi: https://doi.org/10.1016/j.ipm.2021.102724. URL https://www.sciencedirect.com/science/article/pii/S0306457321002089.

[64] D. Tyfield, R. Lave, S. Randalls, and C. Thorpe. *The Routledge Handbook of the Political Economy of Science*. Routledge International Handbooks. Taylor & Francis, 2017. ISBN 978-1-317-41203-8. URL https://books.google.com/books?id=bTQlDwAAQBAJ.

[65] Eric Luis Uhlmann, Charles R. Ebersole, Christopher R. Chartier, Timothy M. Errington, Mallory C. Kidwell, Calvin K. Lai, Randy J. McCarthy, Amy Riegelman, Raphael Silberzahn, and Brian A. Nosek. Scientific Utopia III: Crowdsourcing Science. *Perspectives on Psychological Science*, 14(5):711–733, September 2019. ISSN 1745-6916. doi: 10.1177/1745691619850561. URL https://doi.org/10.1177/1745691619850561. Publisher: SAGE Publications Inc.

[66] Fei-Yue Wang, Wenwen Ding, Xiao Wang, Jon Garibaldi, Siyu Teng, Rudas Imre, and Cristina Olaverri-Monreal. The dao to desci: Ai for free, fair, and responsibility sensitive sciences. *IEEE Intelligent Systems*, 37(2):16–22, 2022. doi: 10.1109/MIS.2022.3167070.

[67] Yizhong Wang, Swaroop Mishra, Pegah Alipoormolabashi, Yeganeh Kordi, Amirreza Mirzaei, Atharva Naik, Arjun Ashok, Arut Selvan Dhanasekaran, Anjana Arunkumar, David Stap, Eshaan Pathak, Giannis Karamanolakis, Haizhi Lai, Ishan Purohit, Ishani Mondal, Jacob Anderson, Kirby Kuznia, Krima Doshi, Kuntal Kumar Pal, Maitreya Patel, Mehrad Moradshahi, Mihir Parmar, Mirali Purohit, Neeraj Varshney, Phani Rohitha Kaza, Pulkit Verma, Ravsehaj Singh Puri, Rushang Karia, Savan Doshi, Shailaja Keyur Sampat, Siddhartha Mishra, Sujan Reddy A, Sumanta Patro, Tanay Dixit, and Xudong Shen. Super-NaturalInstructions: Generalization via declarative instructions on 1600+ NLP tasks. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 5085–5109, Abu Dhabi, United Arab Emirates, December 2022. Association for Computational Linguistics. URL https://aclanthology.org/2022.emnlp-main.340.

[68] K E Weick and K E W Weick. *Sensemaking in Organizations*. Foundations for Organizational Science. SAGE Publications, 1995. ISBN 9780803971776. URL https://books.google.com/books?id=nz1RT-xskeoC.

[69] Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J. G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A. C. 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1):160018, March 2016. ISSN 2052-4463. doi: 10.1038/sdata.2016.18. URL https://www.nature.com/articles/sdata201618. Number: 1 Publisher: Nature Publishing Group.

[70] Amy X. Zhang, Joshua Blum, and David R. Karger. Opportunities and challenges around a tool for social and public web activity tracking. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work Social Computing*, CSCW '16, page 913–925, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450335928. doi: 10.1145/2818048.2819949. URL https://doi.org/10.1145/2818048.2819949.