Semantics-based Framework for Incentivized Research Data Sharing

Kacy Adams, Deborah L. McGuinness, and Oshani Seneviratne
adamsk4@rpi.edu, dlm@cs.rpi.edu, senevo@rpi.edu
Rensselaer Polytechnic Institute
110 8th Street, Troy, NY 12180, USA

Abstract

We present a framework for incentivized research data sharing using an ontology called the *Data Sharing Ontology (DSO)*. The DSO captures the semantics of academic research data sharing and provides an operational specification for data sharing between researchers. The DSO includes a two-part incentive mechanism to confirm citations and reward reproducible research methods. The proposed solution is demonstrated using a dataset-sharing decentralized application use case. The paper's contributions provide a scalable technique for creating, curating, publishing, and consuming web-based, structured, and reusable datasets, including semantically annotated knowledge graphs.

Introduction

The challenges posed by data collection, reuse, and repurposing in scientific studies are multifaceted and go beyond the mere gathering of data. In order to make data useful and effective for research, it must be meticulously organized and annotated. The creation, acquisition, publication, and consumption of large-scale, interlinked structured data corpora have spurred numerous innovative research ideas (Rosenbaum 2010). However, the creation of high-quality, semantics-rich data remains a time-consuming and sometimes laborious process. As a result, researchers often closely guard their datasets (Wilkinson 2008), only making them available to others once they have published academic papers that leverage or introduce the datasets.

Given the importance of data in many research disciplines, there is a growing effort to explore new technologies that can maximize the potential of data, including its ability to be reused, repurposed, and leveraged for economic gain. However, effectively handling datasets and preparing them for reuse requires accurate metadata. For instance, a survey of natural resource scientists, a field that relies heavily on large datasets compiled from multiple sources, revealed that respondents faced various issues when receiving datasets, including missing metadata such as data collection methods and aggregation protocols, as well as incomplete datasets and unresponsive data managers (Volk, Lucero, and Barnas 2014). Intellectual property issues and confidentiality are cited as one of the top reasons health science researchers did not share their data (Alice Meadows 2014). Researchers in less privileged countries may be afraid of getting scooped by more resource-rich institutions (Bezuidenhout and Chakauya 2018). Sometimes researchers with more insightful perspectives on interpreting data may use others' data to their advantage (Fernandez 2010). These challenges stem from a lack of automated methods and tools to assist researchers with annotating their data and a general lack of incentives to promote data sharing in an equitable manner.

There has been a push from journals and policymakers to require publishing data with the corresponding publication, and researchers understand that we need better policies to incentivize and require data sharing (Fecher, Friesike, and Hebing 2015). However, authors' lack of responsiveness to data requests has led to many journal editors reporting draining experiences dealing with authors who submitted papers under clear data sharing policies but subsequently refused to share those data when contacted by other researchers (Savage and Vickers 2009). There is also a change in perceived willingness to share data over time (Campbell et al. 2002). Even though researchers share data primarily to confirm reproducibility, increase citations, and comply with funding mandates (Melero and Navarro-Molina 2020), incentives beyond a conference acceptance or a citation for researchers come few and far between. One incentive, the use of open data badges[1], has been tested in health and medical research, yet studies are unsure of their effectiveness (Rowhani-Farid, Allen, and Barnett 2017).

Although there are several efforts to incentivize researchers, there has not been anything that provides computable and trackable incentives yet. Furthermore, most data sharing interactions are not documented well. By formalizing such a process into a decentralized and automated manner, we begin to eradicate malpractices within research as we reward more robust, well-founded research while democratizing data sharing without relying on a central data repository or publisher. By creating a decentralized history of the data life cycle as it changes hands and finds new use cases, we begin a new expectation within research that investigators should share their data and that others should use it honestly and fairly. As shown in our use case, we enable this expectation via our incentives, the peer review score, and the SCIENCE index we developed.

[1] https://osf.io/tvyxz/wiki/home

Contributions

Our work combines two novel ideas.

First, we capture semantics in peer-to-peer scientific data sharing and reuse through an open-source ontology. We focus on providing a data sharing outlet that facilitates the reproducibility of the utilized research methods and confirms the likelihood of dataset citations. We present the *Data Sharing Ontology (DSO)*, which describes a model for data sharing between researchers who may not have worked together in the past but may mutually benefit from sharing their research datasets. The DSO includes a two-part incentive mechanism to confirm citations and reward reproducible research methods.

Second, we provide an operational specification for decentralized application development for academic research data sharing, guided by the semantic resources developed as part of the first step. We then demonstrate the use of DSO in an exemplary dataset-sharing decentralized application. Our proposed solution will ultimately result in scalable techniques for creating, curating, publishing, and consuming large, web-based, structured, reusable datasets, including semantically annotated knowledge graphs.

Data Sharing Ontology (DSO)

DSO intends to provide a semantic model to power a decentralized application to handle, confirm, and incentivize research data sharing. The ontology uses simple logic such that it could be easily handled by smart contracts (Buterin and others 2014) in decentralized applications. The DSO helps us to capture two categories of semantics. The former is easily accessible, well-known statistics within academic publishing, including the number of publications, citations, and collaborators a researcher has had over their career. The latter category contains less well-known statistics often not considered in the popular academic reputation indices such as the h-index. Such statistics include the quality of research methods, the openness of the research and its artifacts, and the ability to collaborate with others in their field. In capturing all of these, we can more accurately estimate and portray the quality of a researcher's reputation and overall career. In Table 1, we list the ontology prefixes we use to describe objects and their properties in this paper. DSO utilizes properties from the SemanticScience Integrated Ontology (SIO) (Dumontier et al. 2014) and the Provenance Ontology (PROV) (Lebo et al. 2013).

Prefix	Ontology	URI
sio	SemanticScience Integrated Ontology	http://semanticscience.org/ontology/sio.owl
prov	Provenance Ontology	http://www.w3.org/ns/prov-o#
dso	Data Sharing Ontology	http://www.purl.org/rpi/dso#

Table 1: List of Ontologies used in the DSO

Ontology Composition

DSO focuses on a single entity, the collaboration event (`dso:collaborationEvent`). Each collaboration event is between two sets of researchers (`dso:researcher`). The first is a data seeker, referenced by (`dso:hasDataSeeker`), who is seeking a research artifact (`dso:researchArtifact`). The other is a data sharer, referenced by (`dso:hasDataSharer`), who is the owner of the specific research artifact. The collaboration event then follows a set of enumerated states within a custom datatype, (`dso:collaborationState`), beginning with *'Awaiting Data Request'* and ends with *'Contract Destroyed'*. Each participant has a set of activities (`sio:Activity`) that triggers the collaboration event to move to its next state. The states and activities are shown in Figure 1.

Associated Entities

Several of the different activities of the participants require the submission of data. We capture semantics surrounding these entities by defining fields of information that must be included in the dataset transfer operation. These definitions aim to effectively capture these entities while utilizing a decentralized storage system such as the Inter-Planetary File System (IPFS)[2] (Benet 2014).

A very important piece of data sharing is the data request. Many different policies and governing bodies outline data requests in different ways. We decompose the data request into three elements. This includes (1) the group members on the project (who will have access to the data), (2) the affiliated institutions of those involved, and (3) a brief natural language description of the intended use of the data. We believe these elements are vital and non-negotiable components in a data request while allowing additional fields as protocol designers build their system around specific data usage and sharing policies.

We describe the *research artifact* and *publication* via standard metadata methods. As seen in Figure 2 in yellow, the non-published artifact, such as a dataset, is described by its owner, format, and natural language description. In purple, we have a standard publication. Within the protocol, these represent all publications that reuse a given research artifact and are reported to us by the reusing participant. Such a publication would be described by its creator and its unique digital object identifier (DOI). Further within the protocol, we reuse `sio:cites` to point to the research artifact(s) that the publication reuses.

Peer Review Score

The *peer review score* (`dso:haspeerreviewscore`) is a simple model designed to reward honest data sharing and reward reproducible research methods through peer verification while also providing a metric by which data seekers and data sharers can make future decisions on whom to share data with. This public review system will incentivize strong research methods and fair data reuse. The peer review score will capture any deviations from the original agreement (entered when requesting the dataset). These will affect a researcher's future ability to find datasets for reuse within the protocol.

After each completion of a *collaboration event*, participating researchers provide a review (`dso:review`) of the data sharing party through the invocation of a function in the

[2] IPFS is a decentralized, free, content-based storage system.

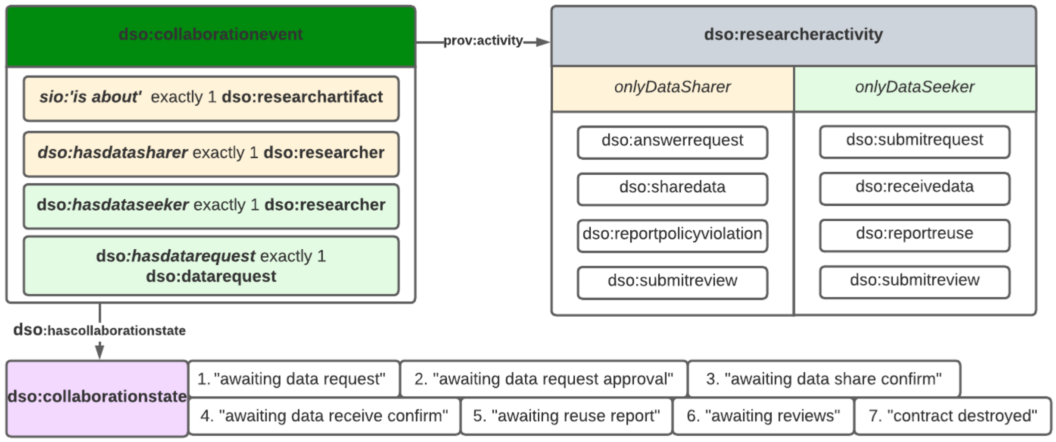

Figure 1: *Collaboration Event* and Associated States & Activities

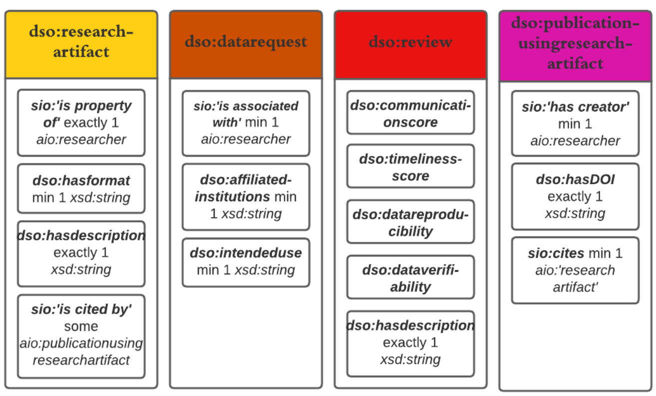

Figure 2: The *Collaboration Event*'s Associated Entities & Information Fields

smart contract. Each review is about (`sio:'is about'`) a specific researcher and is provided by some researcher (using the `sio:'has provider'` property). In Figure 2, the red box outlines the fields of the review as data properties pointing to scores on a scale of 1 through 5. These provide ample room for system designers to expand on these properties to match the space they are building for. A review further allows for a natural language description to give other researchers more context into the provided review. Finally, each review about a researcher is aggregated into a public *peer review score*. A lookup function in the smart contract would enable other researchers to examine someone's peer review score before a collaboration activity.

SCIENCE Index

The SCIENCE index aims to give researchers a standard by which they can compare each other's reputations with respect to research dataset sharing. They can use the index to help make a well-educated decision on whether or not to share their hard-earned research datasets and collaborate with another researcher. The property (`dso:hasSCIENCEindex`) represents a researcher's contributions to science in the form of publications and other research artifacts. The SCIENCE index encompasses an expressive, provenance-centric approach, and it is a recursive acronym for: (i) SCIENCE (ii) Capability-based (iii) Intention-centric (iv) Experiment-oriented (v) Networked (vi) Collaborative (vii) Expression

We measure this SCIENCE metric via statistics such as publication count, citation count, and h-index, as well as a researcher's data sharing history within the protocol. The idea of such an index is to provide a standard for researchers to compare each other when seeking or sharing data. The index is detailed in our previous work (Adams et al. 2023). This paper details the use of a linear model for ranking researchers in comparison to their peers. Based on researchers of a similar career length, we determine an expected h-index and compare this to a researcher's actual h-index to produce the SCIENCE-index. The model provides extensibility to include further parameters to augment the index, which, in our case, would include past data sharing history.

The h-index roughly measures the number of *impactful* publications a researcher has produced. While there are several similar indices that address various inconsistencies in the h-index (Bihari, Tripathi, and Deepak 2021; Waltman and Van Eck 2012), and alternatives such as the proposed "Success Index" (Franceschini et al. 2012), none of them address the impact of dataset sharing in scientific research. We believe the SCIENCE-index, proposed in our prior work would serve as a powerful incentive mechanism for our data sharing protocol.

The `dso:researcher` captures several statistics and identifiers for retrieving these statistics. Specifically, the OrcID[3] and SemanticScholarID[4] allow the protocol to gather a diverse number of statistics via open APIs and standards and calculate the SCIENCE-index in an automated manner.

[3] https://orcid.org

[4] https://www.semanticscholar.org

We include a policy violation activity (`dso:reportpolicyviolation`) for data sharers, allowing them to report a data seeker's data misuse. We anticipate in the future that these policy violations, once confirmed, will impose a penalty on the violating researcher's SCIENCE index, utilizing the index as both an incentive and a disincentive.

Use Case

To demonstrate the general utility of the DSO geared towards peer-to-peer scientific research data sharing in biomedical research that has immensely benefited from advances in semantic web technologies, we present a user story titled "the New Investigator's Dilemma." We utilize this story in Table 2 as a means for framing our competency questions within a realistic example to help readers understand the intention of the DSO.

The New Investigator's Dilemma:

We have two participants, a new investigator and a data seeker. The new investigator, Steven, is a young researcher seeking to build their reputation through important and impactful scientific work. Steven has a large number of citations to a small number of publications. Steven's most recent work includes a well-cultivated dataset representing the single-nucleotide polymorphism (SNP) profiles of individuals with lung cancer, which they understand would be coveted in the biomedical community because of rising, novel work regarding personalized gene therapy treatments utilizing SNP profiles. With the proper incentives, such as citation, recognition, and reputability, Steven would be willing to share the dataset. The data seeker, whom we will refer to as Jeff, is an established researcher seeking to surpass the data creation stage in his next project. We do not know if Jeff means to comply with any policy surrounding the transaction. Steven has the option to share the dataset since it is unknown whether Jeff will participate dishonestly (violate data use policies) or honestly (with proper citations to Steven's dataset, etc.).

Both participants have joined *Sharing-Science*, a new decentralized application leveraging the DSO. Jeff has come across Steven's dataset and is interested in the applications of the individual SNP profiles in the set to other types of cancer. Jeff is concerned by Steven's low SCIENCE index, but they can see that it is because of Steven's low number of publications, as Steven is an early-career researcher. Jeff's concerns are mitigated by Steven's peer review score, which includes several rave reviews of other datasets Steven has shared. Thus, Jeff requests the SNP dataset. Steven, seeing Jeff's intended use in the data request, knowing that Jeff will be penalized if they do not properly cite the dataset, approves the request and shares it with Jeff. Jeff has some questions upon attempting to use the data while also finding the data to be incomplete. Jeff is unable to contact Steven, who is overloaded with other work. Jeff eventually publishes research utilizing the dataset, citing the dataset to avoid penalty. This is confirmed by the protocol, rewarding Steven for his contribution. However, the peer review system documented the lack of communication and issues with the dataset, as Jeff brings up several concerns in their review of the new investigator.

Ontology Development Methodology

The DSO was engineered using a competency question-driven ontology development methodology (Ren et al. 2014). We present a subset of the competency questions used below, describing the coverage of the ontology as a data sharing and handling protocol and the ability of the ontology as an incentive mechanism.

Q1. What does the data seeker intend to use the data for?

Q2. Has the data exchanged hands between the collaborating researchers?

Q3. Which publication by the data seeker uses the shared dataset?

Q4. How many researchers have cited a dataset listed on the protocol?

Q5. Why has a researcher left a specific review?

Q6. What factors are included in the SCIENCE index?

With real-time instances, we must be able to access answers to our first three questions. All stakeholders in a collaboration, as well as the corresponding smart contracts, must be aware of the state of the data request and its specifics (**Q1**), the state of the dataset itself (i.e., has it been physically given to the data seeker?) (**Q2**), and the state of the data seekers reuse of the data (**Q3**).

The latter three questions describe the coverage of the DSO as an incentive scheme. A participating researcher may want to determine how many other researchers have gained access to and cited their dataset (**Q4**). This is an essential incentive as institutions could use it in hiring decisions or funding bodies in grant decisions. A data sharer may seek to know why a poor review has been given to them or one of their shared datasets (**Q5**). This will facilitate reproducible data practices, as a public review will affect other researchers' decisions to request and use the data sharer's datasets. Similarly, the public SCIENCE index, a representation of a researcher's career and data sharing practices (**Q6**), will incentivize good science as it will affect who chooses to share or not share data with whom.

We have provided a sample SPARQL query in Listing 1 that addresses **Q1**, and in Table 2 provide the use case specific competency questions and answer. The full SPARQL queries for our six competency questions are available on our resource website at https://sharing-science.github.io/data-sharing-ontology.

Listing 1: A SPARQL query that retrieves intended use of a data seeker for a given `?collabEvent` instance of `dso:collaborationEvent`

```
PREFIX sio:<http://semanticscience.org/resource/>
PREFIX dso:<http://www.purl.org/rpi/dso#>
SELECT ?intendedUse
WHERE {
  ?collabEvent dso:hasDataRequest ?request .
  ?request dso:intendedUse ?intendedUse . }
```

Question	RDF Characterization	Answer
Q1. What does Jeff intend to use Steven's SNP dataset for?	`dso:dataRequest dso:intendedUse xsd:string`	Jeff intends to use the SNP profiles towards personalized drug therapy research for other types of cancer
Q2. Has Steven physically given the dataset to Jeff?	`dso:collaborationEvent prov:activity dso:shareData`	Steven has shared the dataset with Jeff if the collaboration has an instance of *researcher activity* `shareData`
Q3. Which publication by Jeff used the shared dataset?	`dso:researchArtifact sio:'is cited by' some dso:ResearchArtifactReuse`	Steven's research Artifact is cited by the new publication from Jeff
Q4. How many researchers have cited Steven's SNP dataset before?	count(`dso:researchArtifact sio:'is cited by' dso:ResearchArtifactReuse`)	Four citations from other researchers to the SNP dataset
Q5. Why has Jeff given Steven a poor communication score in his review?	`dso:review dso:communicationScore xsd:integer`	Jeff iterates that Steven was non-communicative during the process
Q6. What factors go into Steven's SCIENCE index, and why is it low?	`dso:researcher sio:'has attribute' dso:hIndex`	Jeff identifies that Steven's SCIENCE-index is low due to a low h-index

Table 2: Use Case Specific Competency Questions (implemented as SPARQL queries), the RDF Triples Returned, and the Natural Language Answer

Related Work

The Data Use Ontology (DUO) (Lawson et al. 2021) presents a set of permissions describing the dataset as general use, health/biomedical use, or disease-specific, a set of modifiers describing attributes of the dataset, e.g., genetic use only, clinical care use, ethical approval required, and a set of semantics to describe the ancestry of the data. Along with their set of specifications for dataset annotations, they capture further semantics describing data use consent, dataset discovery, and data access requests. The DUO captures the semantics of centralized peer-to-peer data sharing but lacks the malleability to survive in a decentralized domain. COEUS (Lopes and Oliveira 2012) gives us a "semantic web in a box" aimed at streamlining ontology-driven application processes. The COEUS platform provides a complete skeleton for rapid interoperability development focused on REST services, a SPARQL endpoint, and Linked Data publication. COEUS aids the development from semantic representation to application, closing the gap between semantic knowledge graphs and their intended use. However, COEUS focuses on centralized application development and does not provide incentives via a mechanism such as the SCIENCE index proposed in this paper.

Coelho et al. (Coelho et al. 2020) present an application built on Hyperledger Fabric (Androulaki et al. 2018), a permissioned blockchain. Their application, known as the E-SECO platform, allows distributed, collaborating researchers to create projects that are handled and tracked by a permissioned blockchain. Each step of the different workflows of academic research is then tracked and logged by the application, allowing researchers to work more efficiently. Other attempts to address many issues in academic publishing, such as predatory publishing, reputation management, and transparent peer-review processes, include the works by Novotny et al. (Novotny et al. 2018), where they present uses for private and public blockchains in academic publishing, describing how they can provide trust and facilitate collaboration without needing a centralized authority. Along the same line, Mackey et al. (Mackey et al. 2019) build a framework for blockchain-based scientific publishing. The framework is built around a shared governance model known as the "Democratic Autonomous Organization" (DAO) made up of validated members representing the different roles of the academic publishing process. The framework handles the different steps of submission, peer review, and editorial decisions that go into the publication workflow in an attempt to increase the transparency of academic publishing via blockchain while maintaining the commonly accepted academic publishing workflow. With the rise of DAOs, which are essentially self-governed, blockchain-based democracies, VitaDAO (DeFrancesco and Klevecz 2022) became the first DAO to mint a Non-fungible Token (NFT) representing the entire legal intellectual property (IP) rights of a biopharma research project. All of the stakeholders of VitaDAO now collectively own this IP and make decisions democratically for the project's future, as well as vote on future funding decisions for other projects. This process paves the way for virtualized IP (Golato 2021), a concept directly applicable to data sharing, signifying that academic research datasets and their corresponding property metadata have a future as virtualized assets.

Discussion and Future Work

Scientific research is not the most efficient in its competitive nature, and we have seen several attempts to increase its efficiency in collaborative workflows (Coelho et al. 2020) and academic publishing governance (Mackey et al. 2019; Novotny et al. 2018). To augment these efforts and realize the potential of existing data while pushing science further, we designed and presented an OWL ontology called the

DSO, modeling a decentralized application that incentivizes peer-to-peer scientific data sharing.

Handling peer-to-peer data sharing is a complex process with several moving parts between policy, participants, and incentives. The DSO helps capture this process with rich semantics and enhanced expressivity, bridging the gap between the data sharing process and its representation within an application. The design flow of smart contracts in a decentralized data sharing protocol includes contracts that handle user roles, data use policies, incentives, and the data sharing events themselves. We began with the DSO contracts and continued designing smart contracts to handle user roles and various use case policies.

We acknowledge the importance of research ethics considerations in data sharing, particularly in the domain of biomedical research. We examined and considered several of the regulatory guidelines and requirements set forth by Research Ethics Committees in designing our proposed protocol. For example, we have explored work in implementing N3C Covid-19 data sharing policies in smart contracts. It is crucial to note that our framework is not intended to undermine any data mandates or violate ethical principles. Rather, we aim to provide a framework that empowers researchers to share their data in a secure, transparent, and traceable manner. Ultimately, the decision to share data remains with the users involved, and we trust that they will abide by the relevant policies and regulations. We believe that our proposed protocol has the potential to promote greater transparency and collaboration in the scientific community while maintaining the highest standards of ethical conduct.

Trust is a crucial aspect of data sharing, and it can be challenging to establish when researchers are working with unfamiliar colleagues or in a highly competitive environment. The peer review score is a means to mitigate this challenge by providing an objective measure of a researcher's reliability, responsiveness, and ability to work collaboratively. By providing a public rating system, the peer review score incentivizes researchers to act professionally and ethically while sharing data, knowing that their reputation is on the line. This then allows researchers to make more informed decisions about potential collaborators, promoting greater transparency and accountability in the research process. We believe that the peer review score can help establish trust between researchers and foster a culture of openness and collaboration, ultimately leading to better scientific outcomes. However, we also acknowledge that there may be limitations to this approach, and further research is needed to evaluate its effectiveness.

The DSO is the first step toward a new type of data ecosystem, but it does not come without challenges. Much work is needed on the implementation and infrastructure for the proposed decentralized application. The SCIENCE index needs to handle extensive statistical calculations and reach external sources, which can be done via a blockchain oracle service such as Chainlink (Breidenbach et al. 2021), but more challenges come in storing and authenticating the data. Further, we do not intend to handle any datasets within the decentralized application or on-chain. Instead, we hope to provide researchers with a simple, packaged solution for storing datasets safely in a decentralized manner using services such as IPFS. Finally, we plan to conduct consistency and completeness checks of the DSO going beyond the biomedical research data sharing domain, as the scientific data sharing needs will change according to the research domain.

Conclusion

In many research fields, the slow and rigorous process of data collection often results in years of work to produce conclusive findings. This inefficiency can lead to the reproduction of similar datasets, leaving datasets with multiple implications unexplored. Additionally, the lack of open datasets forces researchers to collect data themselves, often duplicating the work of others. Thus, there is a critical need for a secure and scalable data sharing ecosystem that incentivizes and rewards collaborative efforts between researchers.

Our solution, the DSO, captures research data sharing activities through smart contracts and includes a novel incentive scheme, the SCIENCE index, that rewards those who share, reuse, and repurpose data. Our approach leverages metrics such as publication count, citation count, and h-index to output a useful metric that incentivizes data sharing in a decentralized manner. These characteristics were identified by a variety of surveys (Volk, Lucero, and Barnas 2014; Melero and Navarro-Molina 2020; Fecher, Friesike, and Hebing 2015) across different fields as the motivating factors of why researchers share data. Through our use case implementation and ontology modeling, we have incorporated terms for decentralized scientific data sharing applications that enable tracking of data sharing, contribution statistics, and collaboration events between researchers. Through a competency question-based evaluation, we illustrated how applications that leverage the DSO could query to track data sharing, contribution statistics, and the state of a collaboration event between researchers. In conclusion, our work seeks to address the problem of data reproduction and inefficient research by promoting incentivized data sharing on the web. Our approach creates a more open, fair, and reproducible knowledge ecosystem that is not reliant on previously established trust relationships, leading to a more efficient research process, and thus accelerating scientific progress.

Resources : We contribute the DSO, the set of formalized SPARQL queries, and the smart contracts implementing the "Sharing-Science" application deployed on an Ethereum test network as publicly available artifacts via persistent URLs. The ontology, documentation and our decentralized data sharing protocol smart contracts have been made available as open-source artifacts under the Apache 2.0 license. We maintain open-source GitHub repositories for all our artifacts.

Project Website: *https://sharing-science.github.io/data-sharing-ontology*

Ontology: *http://www.purl.org/rpi/dso*

References

Adams, K.; Spadea, F.; Flynn, C.; and Seneviratne, O. 2023. Assessing scientific contributions in data sharing spaces.

Alice Meadows. 2014. To Share or not to Share? That is the (Research Data) Question...

Androulaki, E.; Barger, A.; Bortnikov, V.; Cachin, C.; Christidis, K.; De Caro, A.; Enyeart, D.; Ferris, C.; Laventman, G.; Manevich, Y.; et al. 2018. Hyperledger fabric: a distributed operating system for permissioned blockchains. In *Proceedings of the thirteenth EuroSys conference*, 1–15. https://doi.org/10.1145/3190508.3190538: ACM.

Benet, J. 2014. IPFS-content addressed, versioned, p2p file system.

Bezuidenhout, L., and Chakauya, E. 2018. Hidden concerns of sharing research data by low/middle-income country scientists. *Global Bioethics* 29(1):39–54.

Bihari, A.; Tripathi, S.; and Deepak, A. 2021. A review on h-index and its alternative indices. *Journal of Information Science* 01655515211014478.

Breidenbach, L.; Cachin, C.; Chan, B.; Coventry, A.; Ellis, S.; Juels, A.; Koushanfar, F.; Miller, A.; Magauran, B.; Moroz, D.; et al. 2021. Chainlink 2.0: Next Steps in the Evolution of Decentralized Oracle Networks.

Buterin, V., et al. 2014. A next-generation smart contract and decentralized application platform.

Campbell, E. G.; Clarridge, B. R.; Gokhale, M.; Birenbaum, L.; Hilgartner, S.; Holtzman, N. A.; and Blumenthal, D. 2002. Data withholding in academic genetics: evidence from a national survey. *jama* 287(4):473–480.

Coelho, R.; Braga, R.; David, J. M. N.; Dantas, M.; Ströele, V.; and Campos, F. 2020. Blockchain for reliability in collaborative scientific workflows on cloud platforms. In *2020 IEEE Symposium on Computers and Communications (ISCC)*, 1–7. https://doi.org/10.1109/ISCC50000.2020.9219729: IEEE.

DeFrancesco, L., and Klevecz, A. 2022. Decentralized investor communities gain traction in biotech. *Nature biotechnology* 40(9):1310–1315.

Dumontier, M.; Baker, C. J.; Baran, J.; Callahan, A.; Chepelev, L.; Cruz-Toledo, J.; Del Rio, N. R.; Duck, G.; Furlong, L. I.; Keath, N.; et al. 2014. The Semanticscience Integrated Ontology (SIO) for biomedical research and knowledge discovery. *Journal of biomedical semantics* 5(1):1–11.

Fecher, B.; Friesike, S.; and Hebing, M. 2015. What drives academic data sharing? *PloS one* 10(2):e0118053.

Fernandez, R. 2010. Barriers to open science: from big business to Watson and Crick.

Franceschini, F.; Galetto, M.; Maisano, D.; and Mastrogiacomo, L. 2012. The success-index: An alternative approach to the h-index for evaluating an individual's research output. *scientometrics* 92(3):621–641.

Golato, T. 2021. Ip-nfts for researchers: A new biomedical funding paradigm.

Lawson, J.; Cabili, M. N.; Kerry, G.; Boughtwood, T.; Thorogood, A.; Alper, P.; Bowers, S. R.; Boyles, R. R.; Brookes, A. J.; Brush, M.; et al. 2021. The Data Use Ontology to streamline responsible access to human biomedical datasets. *Cell genomics* 1(2):100028.

Lebo, T.; Sahoo, S.; McGuinness, D.; Belhajjame, K.; Cheney, J.; Corsar, D.; Garijo, D.; Soiland-Reyes, S.; Zednik, S.; and Zhao, J. 2013. Prov-O: The prov ontology.

Lopes, P., and Oliveira, J. L. 2012. COEUS:"semantic web in a box" for biomedical applications. *Journal of biomedical semantics* 3(1):1–19.

Mackey, T. K.; Shah, N.; Miyachi, K.; Short, J.; and Clauson, K. 2019. A framework proposal for blockchain-based scientific publishing using shared governance. *Frontiers in Blockchain* 2:19.

Melero, R., and Navarro-Molina, C. 2020. Researchers' attitudes and perceptions towards data sharing and data reuse in the field of food science and technology. *Learned Publishing* 33(2):163–179.

Novotny, P.; Zhang, Q.; Hull, R.; Baset, S.; Laredo, J.; Vaculin, R.; Ford, D. L.; and Dillenberger, D. N. 2018. Permissioned blockchain technologies for academic publishing. *Information Services & Use* 38(3):159–171.

Ren, Y.; Parvizi, A.; Mellish, C.; Pan, J. Z.; Deemter, K. v.; and Stevens, R. 2014. Towards competency question-driven ontology authoring. In *European Semantic Web Conference*, 752–767. https://doi.org/10.1007/978-3-319-07443-6_50: Springer.

Rosenbaum, S. 2010. Data governance and stewardship: designing data stewardship entities and advancing data access. *Health services research* 45(5p2):1442–1455.

Rowhani-Farid, A.; Allen, M.; and Barnett, A. G. 2017. What incentives increase data sharing in health and medical research? a systematic review. *Research integrity and peer review* 2(1):1–10.

Savage, C. J., and Vickers, A. J. 2009. Empirical study of data sharing by authors publishing in PLoS journals. *PloS one* 4(9):e7078.

Volk, C. J.; Lucero, Y.; and Barnas, K. 2014. Why is data sharing in collaborative natural resource efforts so hard and what can we do to improve it? *Environmental management* 53(5):883–893.

Waltman, L., and Van Eck, N. J. 2012. The inconsistency of the h-index. *Journal of the American Society for Information Science and Technology* 63(2):406–415.

Wilkinson, M. 2008. Interoperability with Moby 1.0-it's better than sharing your toothbrush! *Nature Precedings* 1–1.