Evidence Graphs: Supporting Transparent and FAIR Computation, with Defeasible Reasoning on Data, Methods and Results

Sadnan Al Manir[1] [0000-0003-4647-3877],
Justin Niestroy[1] [0000-0002-1103-3882],
Maxwell Adam Levinson[1] [0000-0003-0384-8499],
and Timothy Clark[1,2] [0000-0003-4060-7360]*

[1] University of Virginia, Department of Public Health Sciences, Charlottesville VA USA
[2] University of Virginia, School of Data Science, Charlottesville VA USA

*Corresponding author: twc8q@virginia.edu

Abstract.
Introduction: Transparency of computation is a requirement for assessing the validity of computed results and research claims based upon them; and it is essential for access to, assessment, and reuse of computational components. These components may be subject to methodological or other challenges over time. While reference to archived software and/or data is increasingly common in publications, a single machine-interpretable, integrative representation of how results were derived, that supports defeasible reasoning, has been absent.
Methods: We developed the Evidence Graph Ontology, EVI, in OWL 2, with a set of inference rules, to provide deep representations of supporting and challenging evidence for computations, services, software, data, and results, across arbitrarily deep networks of computations, in connected or fully distinct processes. EVI integrates FAIR practices on data and software, with important concepts from provenance models, and argumentation theory. It extends PROV for additional expressiveness, with support for defeasible reasoning. EVI treats any computational result or component of evidence as a defeasible assertion, supported by a DAG of the computations, software, data, and agents that produced it.
Results: We have successfully deployed EVI for very-large-scale predictive analytics on clinical time-series data. Every result may reference its own evidence graph as metadata, which can be extended when subsequent computations are executed.
Discussion: Evidence graphs support transparency and defeasible reasoning on results. They are first-class computational objects, and reference the datasets and software from which they are derived. They support fully transparent computation, with challenge and support propagation. The EVI approach may be extended to include instruments, animal models, and critical experimental reagents.

Keywords: Provenance, Evidence Graphs, Computation, FAIR

1 Introduction

It is now increasingly understood that dramatically enhanced capabilities for generating and analyzing ultra-large datasets, with increasingly sophisticated methods, requires systematic referencing of archived datasets and software as persistent first-class objects. Formal recording of data provenance is a correlate. There is now beginning to be an additional, and necessary, focus on software citation and provenance. Principles and methods for achieving these goals have been defined and are in various stages of transition to practice in the scientific communications ecosystem [1]–[11].

Incorporating these practices has been advocated as a means to improve verifiability, replicability, reproducibility, and reusability of computational results. The goal, which was established as a requirement at the dawn of modern science [12]–[15], is to make the process by which results and the claims they support, transparent, and to allow the methods to be inspected – at least virtually – for adequacy, and if possible, reused, and improved upon in various applications.

We would like to do this for computations, using a clean, integrated approach that does not require "boiling the ocean", and in which artifacts such as provenance records, which benefit the community as a whole, are generated principally as side-effects of normal computational work, and do not place unwanted burdens or requirements on researchers, with which they cannot realistically be expected to comply. Any system which generates such artifacts ought to have other attributes of significant value to researchers.

Of course, all scientific results are provisional. They, and their interpretations, may be challenged by others in the community, and frequently are. A major reason for transparency itself, is to support such reviews and challenges. Results in science are reviewed "by a jury of peers", similarly to adversarial proceedings in law: a case is made, which may be argued for and against, on the basis of evidence. Published "findings" or claims, and indeed computational results, are not facts. They are defeasible assertions [16], which rely upon a chain of evidence as warrants for belief. That makes them part of a chain of argumentation. We treat computations and their provenance as defeasible arguments for results.

2 Methods

Argumentation frameworks [17]–[26] and abstract dialectical frameworks [27], [28] are important sets of tools and concepts developed in the broader AI community, with an extensive literature.

Bipolar Argumentation Frameworks (BAFs) as developed by Cayrol and others [22], [29], [30], allow both supporting and challenging arguments to be asserted and reasoned over in formal models. Cayrol and Lagasquie-Schiex's work on BAFs supplied inspiration for our approach; which has, however, somewhat different semantics from their model, regarding the support relation, and the meaning of an argument.

In BAFs, arguments are opaque, and without internal structure. This extremely abstract treatment derives from Dung's original presentation [24], enhanced to provide explicit bipolarity. If an argument A supports B, it agrees with B, similar to having an ally in a dispute. Therefore, an attack of C upon A, also attacks B.

Our model, in contrast, treats support as it occurs in the scientific literature, as supporting evidence cited by the author of an argument. If C challenges A, a challenge cannot be inferred on the support for A. However, as in Toulmin argumentation, and the work of Verheij [31]–[33], a challenge to A's support undercuts A – it reduces the warrant for belief in A, if we accept the challenge as valid.

Evidence graphs in our approach may be represented abstractly as:

$$EG = \langle A, R+, R- \rangle, \text{ where:} \quad (1)$$
$$A \text{ is a set of representations;}$$
$$R+ \in A \times A \text{ is the set of } \textit{support} \text{ relations;}$$
$$R- \in A \times A \text{ is the set of } \textit{challenge} \text{ relations; and}$$
$$R+ \cap R- = \emptyset$$

In our treatment, Representations are generalized forms of the various types of assertion found in the scientific literature. All Representations have provenance, if only that they have authorship. A Representation may be a statement, or a set of assertions, claims, or declarations, with provenance as its support.

The simplest form of provenance is attribution of an assertion to an *Agent*. A declarative sentence is an argument of this simple form, where a statement's attribution is its simplest supporting evidence. This corresponds to Aristotle's view of argument as having a significant measure of support from ήθος (ethos), the character of the speaker; and relates to our common experience of tending to give more credence to statements from trusted, highly reputed sources. This notion can be extended to results produced by computational agents whose operations are well validated and transparent.

A *Method* or *Material* representation in a publication, provided as provenance, constitutes a set of assertions about what was done to get a result. Software source code is a set of assertions of this form, when used in a provenance description for a computation.

Clearly, source code has a dual nature, in that as a set of instructions, it is a collection of performatives [34], not evaluable as strictly true or false. We distinguish here between software as a set of instructions to a computer, which does something; and software provided as a description of what was done by the computer, i.e. "it ran this code".

We have adapted these notions to computational provenance representation. We base our model on the following key ideas:

- All data, methods descriptions, and results are sets of defeasible assertions.
- The evidence for correctness of any result is the record of its provenance.
- Subsequent research or discussion may challenge results, dataset, or methods.

The EVI ontology is a formal representation of the evidence for any result or claims as an EvidenceGraph, which unifies existing models for both software and data citation and supports machine-based defeasible reasoning.

We first introduced several of the concepts used in EVI in a previous work on Micropublications [35]. EVI simplifies, revises, and adapts many features of the micropublications model to a purely computational "digital commons" environment, where evidence graphs may be generated by a computation service. The computation service we developed, described elsewhere [36], provides affordances to the user by greatly simplifying access to very large-scale data and underlying parallel computation and workflow services. At the same time, it produces and extends evidence graphs transparently, as a side effect.

3 Results

EVI is an extension of W3C PROV [37]–[39], based on argumentation theory, and supports defeasible reasoning about computations, their results, data, and software. It can be extended to incorporate non-computational evidence important to the results of a computation, for example, the specific instrument (manufacturer and catalog number) used to obtain a dataset, or the reagent used in a binding assay.

Evidence Graphs are directed acyclic graphs, DAGs, produced by a service when a computation is run. They are first-class digital objects and may have their own persistent identifiers and be referenced as part of the metadata of any result. They may be arbitrarily deep. We model these using an OWL 2 vocabulary [40] and set of (https://w3id.org/EVI) which provide for propagation of support and challenge relations, with direct supports / challenges distinguished from indirect. A diagram of the classes and relations is provided in **Fig. 1**.

The classes within the dotted-line box in the figure are basic PROV classes: *Entity*, *Agent*, and *Activity*. All relations in the box are subproperties of PROV relations.

The argumentation relations *supports* and *challenges* are introduced on the new class, *Representation*, a subclass of *prov:Entity*. A *Representation* is a *prov:Entity* that *represents* another *Entity*. All digital objects in EVI are considered to be *Representations*; which may contain other *Representations*.

Supports and *challenges* are superproperties of *directlySupports* and *directlyChallenges* (not shown in the Figure). *Supports* is transitive, *directlySupports* is not. *directlySupports* is a superproperty of the relation *usedBy* (inverse of *prov:used*), and of the property *generates* (inverse of generatedBy) so that if a recorded *Activit*y D used a particular *Representation* C as input, and *generates* E as output, then C *directlySupports* D and D *directlySupports* E.

This is shown in Example 1 below, illustrating how the property chain rule evaluates distant support in the graph, where <s> stands for the *supports* property, and <dS> stands for *directlySupports*.

$$(D \langle used \rangle\ C \wedge D \langle generates \rangle\ E) \Rightarrow \tag{2}$$

$$(C \langle dS \rangle\ D \wedge D \langle dS \rangle\ E) \Rightarrow$$

$$(C \langle s \rangle\ D \wedge D \langle s \rangle\ E) \Rightarrow C \langle s \rangle\ E$$

Supports and *directlySupports* properties are distinguished in this way because they serve different purposes in the model. The *directlySupports* property gives us our DAG,

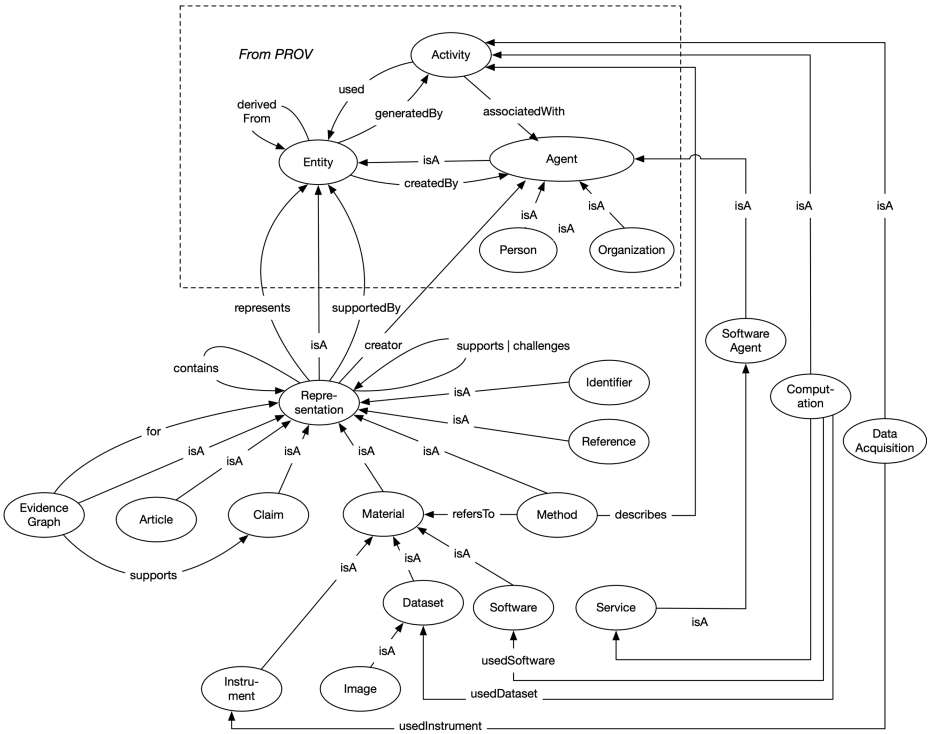

Fig. 1. Classes and relations in the Evidence Graph Ontology, EVI. https://w3id.org/EVI

connecting the various Representations to form an evidence graph. It is simply a generalization over a number of PROV properties. The *supports* property is transitive and allows us to infer distant support relations. By analogy with genealogical trees, *directlySupports* equates roughly to *hasParent* relations; *supports* equates roughly to *hasAncestor* relations.

Articles, *Claims*, *Material* (as used in a typical "Materials and Methods" section), *Methods* (likewise), and *EvidenceGraphs* themselves, are *Representations*. A *Method* describes an *Activity* which may *referTo* a *Material*. The class *Material* is iteratively subclassed into *Instrument, Dataset, Image, Software*, and *Downloads*. *Downloads* are particular distributions of a *Dataset* or *Software*, representing e.g. different formats of a *Dataset*, in the same manner this approach is used in the schema.org class *DataDownload*. An *Agent* in EVI may be a *SoftwareAgent*, *Person*, or *Organization*. A *Service usedBy* a *Computation* is a kind of *SoftwareAgent*. An *Activity* may be a *DataAcquisition* or a *Computation*.

Detailed formal definitions of all these terms, and equivalent schema.org [41] terms where they exist, are provided in the ontology description page at https://w3id.org, which has the latest updates. The OWL 2 vocabulary and its version history are on GitHub, here:

https://github.com/EvidenceGraph/EVI/blob/master/Ontology/versions/v0.2/evi.owl,

and the version we used for this article, is archived on Zenodo [42].

The diagram in **Fig. 1** omits detail relevant to rules and *supports* / *challenges* propagation, for pictorial clarity. The *supports* and *challenges* properties have subproperties *directlySupports*, *indirectlySupports*, *directlyChallenges*, and *indirectlyChallenges*, not shown in the figure. The *supports* property is propagated through the evidence graph via transitivity on *supports*.

$$(\text{A } \textit{directlySupports} \text{ B} \wedge \text{B } \textit{directlySupports} \text{ C}) \Rightarrow \text{A } \textit{supports} \text{ C} \tag{3}$$

Challenges are propagated through a property chain on *challenges* and *supports*.

$$(\text{A } \textit{supports} \text{ B} \wedge \text{X } \textit{challenges} \text{ A}) \Rightarrow \text{X } \textit{challenges} \text{ B} \tag{4}$$

4 Discussion

In developing and validating our approach, we performed a very large-scale time series analysis of ten years of vital signs time series data from the Neonatal Intensive Care Unit (NICU) at the University of Virginia Hospital, collected on 5,957 infants. The NICU time series were analyzed using 80 different algorithms in 11 mathematical families from diverse domains [43]. Computations were performed over a period of 15 months, in various stages, using a FAIR-compliant microservices framework [36], producing an evidence graph of 17,996 nodes [44], queryable from a StardogTM quad store via one of the microservices, and also deposited in the University of Virginia's Dataverse with a *prov:wasDerivedFrom* property associating it with the underlying dataset, also in Dataverse [45].

Fig. 2 illustrates a section of this graph for one of the 5,957 subjects in the NICU analysis. In our microservices framework the EvidenceGraph service stores and retrieves the graphs in a StardogTM quad store, which also performs the inferencing. The PATH query, an extension to SPARQL, generates an *EvidenceGraph* DAG from the root object, using the *directlySupportedBy* abstraction (not shown here) to structure the graph and the *supports* superproperty of its transitive inverse, *supportedBy*, to infer distant evidential support.

As can be seen in the example, every node in the graph has a persistent identifier, based on the ARK system [46] in our implementation. This approach provides the important unification of all available digital evidence for (and potentially, against) a result, by supporting persistent resolution to the cited objects.

One could ask, in a practical sense, if a challenge is made to some leaf node deep in the graph, should it always invalidate the root assertion? The answer to this is that challenges do not invalidate, they present an opposing view. They are ultimately require human judgement as to their validity and strength. We do, however, wish to know about them. This becomes important in Digital Commons environments, and in, for example, meta-analyses reusing prior results. It can have further impact if computational results are properly referenced and tied to textual claims in citation networks, such as those explored by Greenberg [47] and others.

We believe that EVI provides an important generalization of provenance as evidence for correctness, which can be further extended beyond computation to include the other


```
{'@context': {'@vocab': 'http://schema.org/', 'evi': 'http://w3id.org/EVI#'},
 '@id': 'ark:99999/ee797b68-57ac-407d-b995-dbd820283f08',
 '@type': 'Dataset',
 'evi:generatedBy': {'@id': 'ark:99999/bea54627-e761-4c29-b9c0-4ba60f21e3e0',
  '@type': 'evi:Computation',
  'evi:usedDataset': {'@id': 'ark:99999/f94fa455-79dd-4a19-aab8-4e4bef40f12e',
   '@type': 'Dataset',
   'evi:generatedBy': {'@id': 'ark:99999/d24ba632-71cf-4552-a103-9f729850daf8',
    '@type': 'evi:Computation',
    'evi:usedDataset': {'@id': 'ark:99999/bbe4cbef-c073-49c9-baa6-90c52fc2b63d',
     '@type': 'Dataset',
     'author': {'@id': 'https://orcid.org/0000-0002-1103-3882',
      'name': 'Justin Niestroy'},
     'name': 'Raw Time Series Feature Set'},
    'evi:usedSoftware': {'@id': 'ark:99999/2bb5174f-4558-4c69-b339-25c1ebd13c10',
     '@type': 'SoftwareSourceCode',
     'author': {'@id': 'https://orcid.org/0000-0002-1103-3882',
      'name': 'Justin Niestroy'},
     'name': 'Raw Data Analysis Script'},
    'name': 'Computation'},
   'name': 'part-00000-726350cc-f0d3-4b56-a21b-2c44fafab689-c000.csv'},
  'evi:usedSoftware': {'@id': 'ark:99999/8fe92fd8-71b6-491a-a248-c94bcd997406',
   '@type': 'SoftwareSourceCode',
   'author': {'@id': 'https://orcid.org/0000-0002-1103-3882',
    'name': 'Justin Niestroy'},
   'name': 'HCTSA Cluster Script'},
  'name': 'Computation'},
 'name': 'HCTSA Heatmap.png'}
```


Fig. 2. JSON-LD of a portion of the evidence graph for one of the 5,997 infant subjects from clustering step in comparative time-series analysis, adapted from [36].

types of evidence presented in scientific publications, for example by including identifiers such as RRIDs [48] of important experimental reagents and animal models in the protocols from which the data was derived. Our end goal is to be able to provide, with the metadata of any datasets we store in an archive, a link to the corresponding evidence graph. With this practice, any result presented in a publication, with appropriate data citation, ought to be resolvable to its entire evidence graph, and transitively closed to its components.

References

1. Cousijn, H. et al. A data citation roadmap for scientific publishers. Sci Data 5, 180259 (2018).
2. Data Citation Synthesis Group. Joint Declaration of Data Citation Principles. https://doi.org/10.25490/a97f-egyk (2014).
3. Fenner, M. et al. A data citation roadmap for scholarly data repositories. Scientific Data 6, 28 (2019).
4. Groth, P., Cousijn, H., Clark, T. & Goble, C. FAIR Data Reuse – the Path through Data Citation. Data Intelligence 2, 78–86 (2020).
5. Juty, N. et al. Unique, Persistent, Resolvable: Identifiers as the foundation of FAIR. Data Intelligence 2, 30–39 (2020).
6. Katz, D. et al. Recognizing the value of software: a software citation guide [version 2; peer review: 2 approved]. F1000Research 9, (2021).
7. Katz, D. S., Gruenpeter, M. & Honeyman, T. Taking a fresh look at FAIR for research software. Patterns 2, (2021).
8. Smith, A. M., Katz, D. S., Niemeyer, K. E., & FORCE11 Software Citation Working Group. Software citation principles. PeerJ Computer Science 2, e86 (2016).
9. Starr, J. et al. Achieving human and machine accessibility of cited data in scholarly publications. PeerJ Computer Science 1, e1 (2015).
10. Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3, 160018 (2016).
11. Wimalaratne, S. M. et al. Uniform resolution of compact identifiers for biomedical data. Scientific Data 5, 180029 (2018).
12. Dear, P. Revolutionizing the Sciences: European Knowledge and Its Ambitions, 1500-1700. (Princeton University Press, 2009).
13. Holmes, F. L. Argument and Narrative in Scientific Writing. in The Literary Structure of Scientific Argument: Historical Studies (ed. Dear, P.) 224 (University of Pennsylvania Press, 1991).
14. Rossi, P. Philosophy, Technology, and the Arts in the Early Modern Era. (Harper & Row, 1970).
15. Shapin, S. Pump and Circumstance: Robert Boyle's Literary Technology. in The Scientific Revolution (ed. Hellyer, M.) (Blackwell, 2003).
16. Koons, R. Defeasible Reasoning. Stanford Encyclopedia of Philosophy (Spring 2014 Edition) (2013).
17. Bench-Capon, T. J. M. & Dunne, P. E. Argumentation in artificial intelligence. Artificial Intelligence 171, 619–641 (2007).
18. Besnard, P. & Hunter, A. Elements of Argumentation. (MIT Press, 2008).
19. Boella, G., D.M.Gabbay, Van Der Torre, L. & Villata, S. Support in Abstract Argumentation. in Computational Models of Argument (eds. Baroni, P. et al.) (IOS Press, 2010).
20. Brewka, G., Polberg, S. & Woltran, S. Generalizations of Dung Frameworks and Their Role in Formal Argumentation. IEEE Intell. Syst. 29, 30–38 (2014).
21. Carrera, Á. & Iglesias, C. A. A systematic review of argumentation techniques for multi-agent systems research. Artif Intell Rev 44, 509–535 (2015).
22. Cayrol, C. & Lagasquie-Schiex, M.-C. Bipolar Abstract Argumentation Systems. in Argumentation in Artificial Intelligence (eds. Rahwan, I. & Simari, G. R.) (Springer, 2009).
23. Cohen, A., Gottifredi, S., García, A. J. & Simari, G. R. An approach to abstract argumentation with recursive attack and support. Journal of Applied Logic 13, 509–533 (2015).

24. Dung, P. M. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artificial Intelligence 77, 321–357 (1995).
25. Oren, N. & Norman, T. J. Semantics for Evidence-Based Argumentation. 9 (2003).
26. Argumentation in Artificial Intelligence. (Springer, 2009).
27. Brewka, G. & Woltran, S. Abstract Dialectical Frameworks. 10 (2010).
28. Dung, P. M. & Thang, P. M. Representing the semantics of abstract dialectical frameworks based on arguments and attacks. Argument & Computation 9, 249–267 (2018).
29. Cayrol, C. & Lagasquie-Schiex, M.-C. Coalitions of arguments: A tool for handling bipolar argumentation frameworks. International Journal of Intelligent Systems 25, 83–109 (2010).
30. Cayrol, C. & Lagasquie-Schiex, M.-C. Bipolarity in argumentation graphs: Towards a better understanding. International Journal of Approximate Reasoning 54, 876–899 (2013).
31. Toulmin, S. E. The Uses of Argument. (Cambridge University Press, 2003).
32. Verheij, B. The Toulmin Argument Model in Artificial Intelligence. Or: how semi-formal, defeasible argumentation schemes creep into logic. in Argumentation in Artificial Intellgence (eds. Rahwan, I. & Simari, G.) (Springer, 2009).
33. Verheij, B. Evaluating Arguments Based on Toulmin's Scheme. Argumentation 19, 347–371 (2005).
34. Austin, J. L. How to Do Things With Words. (Harvard University Press, 1962).
35. Clark, T., Ciccarese, P. N. & Goble, C. A. Micropublications: a semantic model for claims, evidence, arguments and annotations in biomedical communications. J Biomed Semantics 5, 28 (2014).
36. Levinson, M. A. et al. FAIRSCAPE: A Framework for FAIR and Reproducible Biomedical Analytics. 2020.08.10.244947 http://biorxiv.org/lookup/doi/10.1101/2020.08.10.244947 (2020) doi:10.1101/2020.08.10.244947.
37. Gil, Y. et al. PROV Model Primer: W3C Working Group Note 30 April 2013. (2013).
38. Lebo, T. et al. PROV-O: The PROV Ontology W3C Recommendation 30 April 2013. (2013).
39. Moreau, L. et al. PROV-DM: The PROV Data Model: W3C Recommendation 30 April 2013. http://www.w3.org/TR/prov-dm/ (2013).
40. OWL 2 Working Group. OWL 2 Web Ontology Language: W3C Recommendation 27 October 2009. http://www.w3.org/TR/owl-overview/ (2009).
41. Guha, R. V., Brickley, D. & Macbeth, S. Schema.org: evolution of structured data on the web. Communications of the ACM 59, 44–51 (2016).
42. Manir, S. A., Niestroy, J., Levinson, M. & Clark, T. EVI: The Evidence Graph Ontology, OWL 2 Vocabulary. (University of Virginia, 2021).
43. Niestroy, J. et al. Discovery of signatures of fatal neonatal illness in vital signs using highly comparative time-series analysis. (2021).
44. Niestroy, J. et al. Evidence for: Discovery of signatures of fatal neonatal illness in vital signs using highly comparative time-series analysis. (2021).
45. Niestroy, J. et al. Replication Data for: Discovery of signatures of fatal neonatal illness in vital signs using highly comparative time-series analysis, V2. (2021).
46. Kunze, J. & Rodgers, R. The ARK Identifier Scheme. (2008).
47. Greenberg, S. A. Understanding belief using citation networks. Journal of Evaluation in Clinical Practice 17, 389–393 (2011).
48. Bandrowski, A. E. & Martone, M. E. RRIDs: A Simple Step toward Improving Reproducibility through Rigor and Transparency of Experimental Methods. Neuron 90, 434–436 (2016).