JOURNAL OF
BIOMEDICAL SEMANTICS

RESEARCH   **Open Access**

Micropublications: a semantic model for claims, evidence, arguments and annotations in biomedical communications

Tim Clark[1,2,3*], Paolo N Ciccarese[1,2] and Carole A Goble[3]

Abstract

Background: Scientific publications are documentary representations of defeasible arguments, supported by data and repeatable methods. They are the essential mediating artifacts in the ecosystem of scientific communications. The institutional "goal" of science is publishing results. The linear document publication format, dating from 1665, has survived transition to the Web.

Intractable publication volumes; the difficulty of verifying evidence; and observed problems in evidence and citation chains suggest a need for a web-friendly and machine-tractable model of scientific publications. This model should support: digital summarization, evidence examination, challenge, verification and remix, and incremental adoption. Such a model must be capable of expressing a broad spectrum of representational complexity, ranging from minimal to maximal forms.

Results: The micropublications semantic model of scientific argument and evidence provides these features. Micropublications support natural language statements; data; methods and materials specifications; discussion and commentary; challenge and disagreement; as well as allowing many kinds of statement formalization.

The minimal form of a micropublication is a statement with its attribution. The maximal form is a statement with its complete supporting argument, consisting of all relevant evidence, interpretations, discussion and challenges brought forward in support of or opposition to it. Micropublications may be formalized and serialized in multiple ways, including in RDF. They may be added to publications as stand-off metadata.

An OWL 2 vocabulary for micropublications is available at http://purl.org/mp. A discussion of this vocabulary along with RDF examples from the case studies, appears as *OWL Vocabulary and RDF Examples* in Additional file 1.

Conclusion: Micropublications, because they model evidence and allow qualified, nuanced assertions, can play essential roles in the scientific communications ecosystem in places where simpler, formalized and purely statement-based models, such as the nanopublications model, will not be sufficient. At the same time they will add significant value to, and are intentionally compatible with, statement-based formalizations.

We suggest that micropublications, generated by useful software tools supporting such activities as writing, editing, reviewing, and discussion, will be of great value in improving the quality and tractability of biomedical communications.

Keywords: Argumentation, Annotation, Data citation, Digital abstract, Scientific discourse, Scientific evidence, Methods citation, Research reproducibility, Nanopublications

* Correspondence: tim_clark@harvard.edu
[1]Department of Neurology, Massachusetts General Hospital, 55 Fruit Street, Boston, MA 02114, USA
[2]Harvard Medical School, 25 Shattuck Street, Boston, MA 02115, USA
Full list of author information is available at the end of the article

Introduction

During the past two decades the ecosystem of biomedical publications has moved from a print-based to a mainly Web-based model. However, this transition brings with it many new problems, in the context of an exponentially increasing, intractable volume of publications [1,2]; of systemic problems relating to valid (or invalid) citation of scientific evidence [3,4]; rising levels of article retractions [5,6] and scientific misconduct [7]; of uncertain reproducibility and re-usability of results in therapeutic development [8], and lack of transparency in research publication [9]. While we now have rapid access to much of the world's biomedical literature, our methods to organize, verify, assess, combine and absorb this information in a comprehensive way, and to move discussion and annotation activities through the ecosystem efficiently, remain disappointing.

Computational methods previously proposed as solutions include ontologies [10]; text mining [2,11,12]; databases [13]; knowledgebases [14]; visualization [15]; new forms of publishing [16]; digitial abstracting [1]; semantic annotating [17]; and combinations of these approaches. However, we lack a comprehensive means to orchestrate these methods. We propose to accomplish this with a layered metadata model of scientific argumentation and evidence.

Such a common metadata representation of scientific claims, argument, evidence and annotation in biomedicine should serve as an integrating point for the original publication, subsequent annotations, and all other computational methods, supporting a single framework for activities in the nine point cycle of authoring-publishing-consumption-reuse we discuss in the section on Use Cases. This cycle can be thought of as an information value chain in science. This means that each set of disparately motivated and rewarded activities, carried out by various actors, creates and passes along value to the next, which consumes this value-added product as an input. A metadata representation to support this value chain would need to:

- serve as a common Web-friendly nucleus for value-addition and extraction across the biomedical communications ecosystem: understood, operated upon and exchanged by humans and by computers, as supplements to the linear documents they characterize;
- enable more powerful use and sharing of information in biomedicine, particularly through integration and mashup to provide the most relevant views for any social unit of researchers;
- enable the addition of value to the content while providing a detailed provenance of what was done;
- support computational processing in a way that complete papers in un-augmented linear natural language cannot yet integrate well with existing linear textual representations.

This paper introduces the micropublications semantic metadata model. The micropublications model is adapted to the Web, and designed for (a) representing the key arguments and evidence in scientific articles, and (b) supporting the "layering" of annotations and various useful formalizations upon the full text paper.

This model responds to the nine use cases we present, in which digital summarization of scientific argumentation with its evidence and methodological support is required. These use cases, for the most part, deal directly with the scientific literature, rather than its processed reflection in curated topical databases. They illustrate how and why currently proposed "statement-based" approaches need richer representation and how this model can play such a role.

In this paper we present

- a Use Case analysis mapped to sets of common activities in the biomedical communications ecosystem, showing the potential value addition and path to implementation of the proposed model for each Use Case;
- a formal model of micropublications;
- illustrative examples instantiating the model for each Use Case;
- notes on an interface to nanopublications and other statement-based formalizations;
- discussion on how the model can support reproducibility and verifiability in research; on implementation in software; and relationship to other work; and
- our Conclusions about the role of this model in next-generation scientific publishing.

We also provide, in three separate files of Additional Material:

1. detailed class, predicate and rule definitions;
2. a proposed Web-friendly representation, using community ontologies, serialized in the W3C Web Ontology Language, with a set of examples in RDF; and
3. a comparison of micropublications to the SWAN model.

Background
Beyond statement-based models

Statement-based models have been proposed as mechanisms for publishing key facts asserted in the scientific literature or in curated databases in a machine processable form. Examples include: Biological Expression Language

(BEL) statements [18]; SWAN, a model for claims and hypotheses in natural language developed for the annotation of scientific hypotheses in Alzheimers Disease (AD) research [14,19-21]; and nanopublications [22-26], which contribute to the Open PHACTS linked data warehouse of pharmacological data [26].

What we mean by "statement-based" is that they confine themselves to modeling statements found in scientific papers or databases, with limited or no presentation of the backing evidence for these statements. Some offer statement backing in the form of other statements in the scientific literature, but none actually has a complete representation of scientific argument including empirical evidence and methods. Of the three examples we mention,

- Nanopublications model only the indicated statement;
- SWAN models a principal statement, or "hypothesis", with supporting statements, or "claims", from the same publication only, and backing references for the supporting statements;
- BEL and SWAN model backing statements from other publications in the literature by citing whole publications, leaving the reader to determine precisely where in the cited document a backing statement actually resides;
- None of these models provide a means to build claim networks of arbitrary depth.
- None of these models provide a means to transitively close claim lineages to underlying empirical evidence – because they do not represent it.

Table 1 compares these three statement-based models.

Figure 1 shows an example of a nanopublication which attempts to express the assertion from Spilman et al. [27] that "inhibition of mTOR by rapamycin can slow or block AD progression in a transgenic mouse model of the disease". Nanopublications distill content as a graph

Table 1 A comparison of SWAN, nanopublications and Biological Expression Language

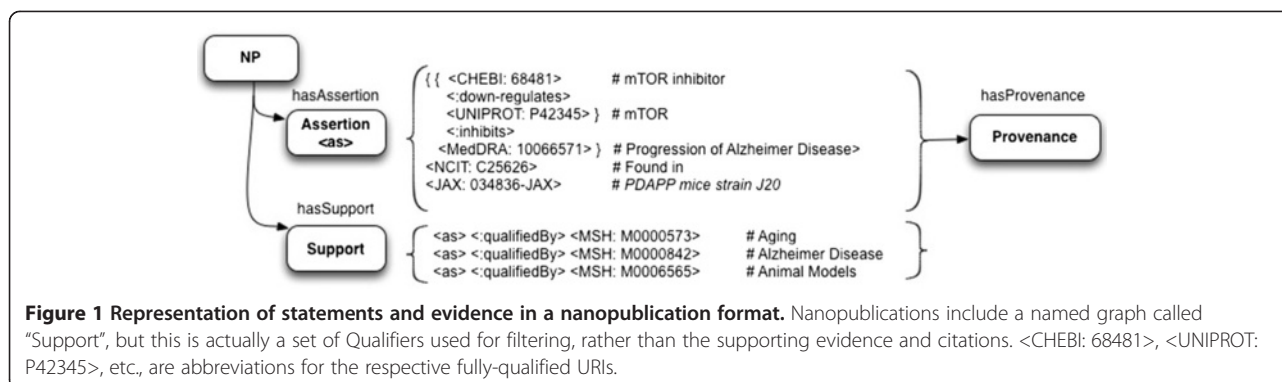|	SWAN	Nanopublications	BEL
First specification	2008 [20]	2010 [24]	2011 [18]
Current specification	2012 [19]	2012 [25]	2011
Statements in	Natural language	Formal language	Formal language
Statement provenance	SWAN-PAV ontology	SWAN-PAV ontology	SWAN-PAV ontology
Backing references from literature	Yes	Yes	Yes
Support claim networks	No	No	No
Direct empirical scientific evidence	No	No	No

of assertions associated with (a) provenance of the article or dataset from whence they came; and (b) a set of terms for indexing and filtering in order to identify auxiliary information in large data sets. Although this last point is represented by a named graph called "Support", this is not intended to represent argumentative support or evidence, but rather descriptive information (cell type, species, etc.) "to enable first pass filtering over large nanopublication sets" [25]. Note that formalization of the np:Assertion is somewhat awkward in this example, and requires multiple level of reification. Yet the np:Assertion is not modelling a markedly complex scientific claim.

The intent of statement-based models is to be relatively simple and useful for specific tasks. In the case of nanopublications, this particular model is currently presented (on a technical level) mainly for data integration across chemical and biological databases. For example, in the current nanopublication guidelines [25], a nanopublication is declared to be "a layer on top of RDF encoded data to provide a standard for the identification of individual scientific assertions within a dataset [which] enables the provenance to be assigned to each assertion and the entire dataset itself". There is no suggestion in the current specification that nanopublications may be applied directly to ordinary scientific articles, nor that they are designed to present primary scientific evidence – although more expansive claims have been made elsewhere in the literature [22,23]. Furthermore, the fact that formalization of assertions is required, is likely an impediment to such direct use. Consequently, a more comprehensive model is needed to be applied successfully across the entire ecosystem of biomedical communications.

The micropublication approach goes beyond statements and their provenance, proposing a richer model in order to account for a more complete and broadly useful view of scientific argument and evidence, beyond that of simple assertions, or assertions supported only by literature references. It is also designed to be readily compatible with assertions coded in BEL or as nanopublications, as these models are considered useful in certain applications and will need to be integrated.

The role and importance of empirical evidence
Empirical evidence is required in scientific publications so that the scientific community may make and debate judgments based on the "interpretation of nature" rather than interpretation of texts [28]. The process of establishing new "facts" and either supplementing or overthrowing old ones, is the central work of biomedical research. Scientific assertions do not become matters of fact until the facts have been established, through judgements made over time in a complex social process. This process includes collective investigation and assessment, and may involve controversy, uncertainty, overthrow of

Figure 1 Representation of statements and evidence in a nanopublication format. Nanopublications include a named graph called "Support", but this is actually a set of Qualifiers used for filtering, rather than the supporting evidence and citations. <CHEBI: 68481>, <UNIPROT: P42345>, etc., are abbreviations for the respective fully-qualified URIs.

settled opinion, and gradual convergence over time on a "current best explanation" [29]. It takes place as researchers present arguments in the professional literature; with supporting observational data, interpretations, and theoretical and methodological context, for evaluation by a "jury of their peers". Once a matter of fact is established, the scientific literature persists as an open documentary record, which from time to time may be challenged and reassessed. Thus, to usefully model facts *in the process of formation*, empirical evidence, as well as formal statements and their provenance, must be a part of our model.

For example, consider the following "nano-publishable" fact given in [23,24]: "mosquitos transmit malaria". This example reflects old science that is not currently under examination or contention, as the role of some species of *Anopheles* as malarial vectors has been well-established for over a century (roughly since the period leading up to Ronald Ross's 1902 Nobel Prize in Medicine). However, previously, in the late nineteenth century, the existence and nature of malarial vectors was an open research question [30]. Open research questions require *presentation of evidence* to establish a *warrant for belief* [31]. Thus, for scientists working on malaria over a century ago, a statement about supposed malarial vectors without supporting empirical evidence, would not have been robust enough to enable evaluation, and thus could not have motivated reasoned belief.

Recent results spotlight concerns with the communication of evidence and its citation. Begley and Ellis recently found that only 11% of research findings they examined from the academic literature could be reproduced in a biopharmaceutical laboratory [8]. Fang et al. reviewed all retractions indexed in PubMed, finding that over two thirds were due to misconduct [7]. Retractions themselves are an increasingly common event [6]. Greenberg conducted a citation network analysis of over 300 publications on a single neuromuscular disorder, and found extensive progressive distortion of citations, to the extent that reviews in reputable journals presented statements as "facts", which were ultimately based on no

evidence at all [3,4]. Simkin and Roychowdhury showed that, in the sample of publications they studied, a majority of scientific citations were merely copied from the reference lists in other publications [32,33]. The increasing interest in direct data citation of datasets, deposited in robust repositories, is another result of this growing concern with the evidence behind assertions in the literature [34].

We have incorporated a number of features in our model to enable presentation of empirical scientific evidence; therefore including data, not just assertions, as information supporting a statement; as well as other required features for scientific discourse.

The importance of natural language

As useful as formal language representations may be, any requirement that statements must only be expressed in formal language such as we find in BEL, nanopublications, and some other approaches, is a potential barrier to adoption in the publication ecosystem. We can expect to encounter scientific claims in their native environment, the biomedical literature, as relatively nuanced *arguments for qualified claims supported by evidence*. This evidence consists of citations to the literature, and novel data with supporting methods. Scientific claims "in the wild" are almost always extensively hedged or qualified, based on recognition of their incomplete or tentative nature [35]. Moreover, ordinary scientific workers present their conclusions in natural language and will continue to do so. Previous experiments such as "structured digital abstracts" have faltered: authors have little incentive to formalize their claims, only to publish them [36,37], and tooling support is poor for those that have that desire.

Consequently the micropublication model must capture the natural language of claims as they appear in the literature. We treat formalization separately as an optional curatorial step. This more comprehensive approach deals with both with scientific statements "in the wild", and with the evidence that supports them, and is also compatible with statement-based formalization patterns such as nanopublications.

Methods

Formalizing scientific publications as arguments

Our model is based on understanding scientific publications as arguments, which present a narrative of experiments or observations, the data obtained, and a reasoned interpretation ("finding") of the data's meaning [38]. Such arguments present a line of reasoning, to a "best" explanation of the data ("abductive reasoning", "inference to the best explanation", "ampliative inference") [39-42] taken in context with the published findings of others in the field.

The determination whether or not a finding is correct, is made over time by the community of the researcher's peers. What claims are considered true, may evolve over time, based on re-examination of evidence and development of new evidence. Assertions may be criticised and refuted. Thus scientific reasoning is *defeasible* [43].

Toulmin's classic model of defeasible reasoning [44], updated by Bart Verheij [45,46], focuses on the internal structure of argument: what the author states, how it is qualified, how the author backs it up, and what other arguments may contradict it. It is a mainstream model of argument in the Artificial Intelligence (AI) community. In our model, we extend both the "support" and the "contradiction" or "rebuttal" part of this model to interargumentation, or argumentation frameworks, a topic with its own extensive literature in AI (see, e.g., [47-53], etc.). The micropublication model is grounded in Toulmin-Verheij; and is consistent with recent work in AI on defeasible argumentation [43,45,48,50,51,54].

Our model provides a framework to support extensively qualified claims in natural language, as generally presented by researchers in their primary publications. Most fundamentally, it adds support relations to claims in the literature, to assist in resolving primary scientific evidence within a "lineage" of assertions. Support relations, structured as graphs, back up assertions with the data, context and methodological evidence which validates them.

Micropublications permit scientific claims to be formulated minimally as any statement with an attribution (basic provenance), and maximally as entire knowledgebases with extensive evidence graphs.

Thus micropublications in their minimal form subsume or encompass statement-based models, while allowing presentation of evidential support for statements and natural language assertions as backing for formalisms. This has significant applicability across the lifecycle of biomedical communications.

Use case analysis

The goal of the micropublication model is to better adapt scientific publications to production and use on the Web, in the context of the new forms either made available or required. The model supports nine main activities (e.g.

authoring, reviewing, etc.) in our analysis, within nine use case requirement sets (e.g. building citable claims).

We begin our analysis of the model's use case applicability by abstracting major activities related to publishing scientific articles, the benchmark of scientific accomplishment [55] within the biomedical communications ecosystem. These are activities in the lifecycle of biomedical communications – part of its knowledge or information value chain [56-58] – to which the model is meant to respond, and which provide context for the model-specific use cases.

Within these we select a set of important but non-exhaustive applications of the model, targeted at responding to specific deficits already identified. For users of scientific publications, these are mostly centered on failures of evidence and reproduciblity [8,55,59,60]; and on intractable volumes of information presented to the scientist [61].

Applications of the model begin with its feature of citable claims (use case 1), supported by evidence (use case 2), from which a robust claims network may be automatically or semi-automatcally constructed and analyzed (use case 3). These first three use cases respond to the identified problems of mishandled, degraded or fictitious citations [3,4,32,62]; and to scientific claims not properly grounded in evidence [3,4]. These ultimately all address the issue of scientific reproducibility [8,55,59,60].

We then provide a use case centered on abstracting single articles (use case 4). This responds to the publication-volume overload issue noted by Cohen and Hunter, and others [1,63], by facilitating useful operations of various computational browsing tools such as [64]; construction of structured claim-evidence representations within reference managers; and as a side effect, enabling the construction of claim networks already mentioned (see above).

Topic-centric claim networks may be equivalent to domain-centric knowledgebasees, if all claims having common meaning, but different wording, can be made functionally equivalent (use case 5). They are more readily computable, if formalized (use case 6) in languages such as Biological Expression Language (BEL) [18]; or nanopublications [24,26,65]. They may be developed under formal curation, for a department or other research enterprise; or as extensions to bibliographic data management by individual scientists.

Publications must be discussed in the biomedical community. This is a part of their validation. Making online discussion part of the permanent record in a compact way related directly to the claim network of an article, constitutes use case 7, and again responds to the critical "filtration" and flagging of invalid or disputed results [66,67]. Use case 7 also responds to the increasing interest in algorithmic annotation using semantic models [2,11,17,68-84] by providing a way for computational
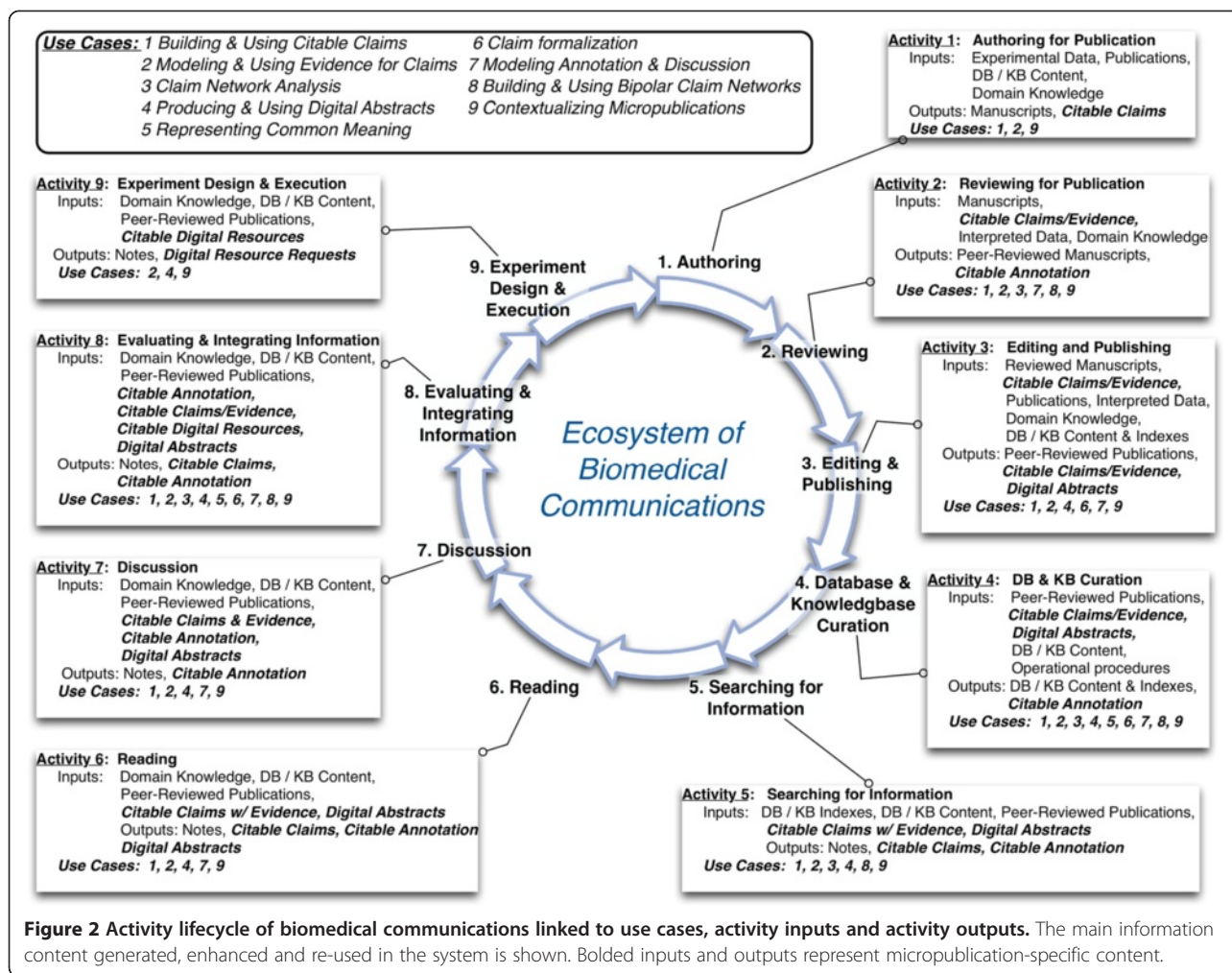
annotation to be combined with argument models. Use case 8 responds to the nature of scientific publications and discussions as arguments, which may agree of disagree with one another, and allows findings from the argumentation theory community [49,54] to be deployed on constructed claim networks.

Lastly, if this model is to be deployed, it must be backward-compatible with the existing communications ecosystem. For this we rely on emerging stand-off web annotation models (use case 9) [85-88].

Table 2 shows the mapping of activities to use cases and Figure 2 shows how each of these use cases, and their inputs and outputs, are situated in an "activities lifecycle", across the biomedical communications ecosystem. This lifecycle is part of a value chain. We will show here how the micropublications model can effectively support information creators and consumers (*tool users*) in this ecosystem, for important unsatisfied use cases.

The use cases in Figure 2, with their motivations and uses, are described in detail below. For each one we show a motivation, use, point of implementation, and comments (if any). By "point of implementation" we mean a practical activity already in place in the ecosystem, in which the model could be implemented, within some new functionality that returns value to a user.

1. **Building and using citable claims**
   Motivation: Simkin and Roychodhury used mathematical techniques to show that scientific authors only read approximately 10%-20% of the papers they cite, and many of the rest are evidently copied from reference lists [32,33]. Greenberg's [3,4] analysis of the distortion and fabrication of claims in the biomedical literature demonstrates why citable claims are necessary. In his analysis, it is straightforward to see how citation distortions may contribute to non-reproducible results in a pharmaceutical context, as reported in [8].

Use: Citable claims are a specific remedy for citation distortion by allowing ready comparison of what is cited, to what the citation is claimed to assert. Implementation: Citable claims may be constructed economically at the point where researchers read and take notes upon, or search for backing for their own assertions in, the domain literature of their field.
Comments: Any scientific statement with an attribution may be formalized as a citable claim using the micropublication model.

2. **Modeling evidence support for claims**
   Motivation: Evidence is the basis for assessment and validation of claims in biomedical (and scientific) argument. Greenberg [3,4] specifically showed how citation lineages may not actually resolve to empirical evidence. Claims ultimately must be based on data, and data must be based on reproducible methods.
   Use: Micropublications may be used to represent experimental evidence supporting claims, as they can represent non-statement artifacts such as reagents, images and other data. This function of the model has multiple roles. It adds additional value to citable claims by indicating what claims are actually backed by direct evidence, and what this evidence is. It also provides the ability to trace the association of claims in the literature to specific methods and data, and vice versa.
   Implementation: As in use case 1, evidence support for claims may be modeled as part of the process of recording bibliographic references. It may also be modeled directly by publishers as supplemental metadata, or by biomedical Web communities as part of a discussion.

3. **Producing a digital abstract of a publication**
   Motivation: Digital abstracts would be extremely useful supplemental metadata. They would be

Table 2 Mapping of Activities to use cases for micropublications

Use case ↓ Activity→	Author	Review	Edit & publish	DB/KB curate	Search for Info	Read	Discuss	Evaluate & integrate	Design & execute experim't
1 - Build and Use Citable Claims	X	X	X	X	X	X	X	X	
2 - Model Evidence for Claims	X	X	X	X	X	X	X	X	X
3 - Prepare Digital Abstracts		X		X	X			X	
4 - Analyze Claim Networks			X	X	X	X	X	X	X
5 - Establish Common Meaning				X				X	
6 - Formalize Claims			X	X				X	
7 - Annotate & Discuss		X	X	X		X	X	X	
8 - Model Bipolar Claim Networks		X		X	X			X	
9 - Context-ualize Model in Documents.	X	X	X	X	X	X	X	X	X

Figure 2 Activity lifecycle of biomedical communications linked to use cases, activity inputs and activity outputs. The main information content generated, enhanced and re-used in the system is shown. Bolded inputs and outputs represent micropublication-specific content.

particularly useful to enable text mining as argued by Gerstein et al. [36].

Use: A micropublication could be used to formalize the supporting attribution, central claims, references to literature, scientific data, materials & methods, annotations, comments, and formalizations, of scientific communications.

Implementation: They could be provided by publishers or by value-add third parties, or created as part of personal or institutional knowledge bases. Mashing-up digital abstracts can be done by third-party applications, and would be one way to deal with intractable publication volumes, by properly summarizing them in a reliable, computable way.

Comments: To enable complete digital abstracting, we define a system of classes for representing biological objects such as reagents, software, datasets and method descriptions, which are not statements in natural language or triples, but are important in documenting the foundational evidence for biomedical claims and arguments, and in making biomedical methods reusable.

4. **Claim network analysis of publication sets**

Motivation: As previously noted, it has been shown that the biomedical literature contains a significant proportion of non-reproducible results. These can be made even more problematic as they are repeatedly cited and transformed in claim networks.

Use: Claim network analysis can be used to determine the origin of, and compare evidence for, individual and contrasting claims in the literature. Particularly when experiments are being designed based on putative findings of a body of prior research; it seems critical to be able to fully assess the entire background of a set of assertions.

Implementation: Micropublications once instantiated, embody individual arguments, which in turn may be composed by resolution of references. This allows us to create extended graphs showing the basis in evidence for claims in the literature, even when they are deeply buried in chains of

citations. Claim lineages are chains of citing/cited claims. Lineage visualization is proposed as a tool for readers and reviewers.

5. **Representing common meaning using similarity groups**

Motivation: Resolution of references often entails finding a claim in a cited document, which is similar to the claim formulated in the citing document. Further, parallel claims, of equivalent or near-equivalent meaning, may arise from different lines of research, without resolution to a common progenitor study.

Use: Using *Similarity groups*, a set of claims may be defined as having "sufficient" closeness in meaning to a representative exemplar, or *Holotype* claim. Their purpose is to allow normalization of diverse sets of statements with essentially the same meaning in the literature, without combinatorial explosion. We term the members of a given equivalence group, *"equivalents"* of one another.

Implementation: Holotypes may be defined (a) when a backing statement for a claim is defined, by choosing one or the other as the holotype, or by defining a new "annotator's version" as a holotype; (b) in a similar way, when a text similarity search on the library of claims detects similar statements in separate claim lineages.

Comments: The *similog-holotype model* is an empirically based model that allows similar claims to be normalized to a common natural language representation, without dropping necessary qualifiers and hedging. *Translations* of claims to formal or other natural languages may also be considered similogs to the translated original, based on (sufficient) equivalence of meaning.

6. **Claim formalization with attribution**

Motivation: Various applications in computing require translation of natural-language claims in the biomedical literature to statements in a formal vocabulary. Biological Expression Language (BEL) [18] and Attempto Controlled English (ACE) [89-91] are examples of claim formalization vocabularies, as are nanopublications.

Use: Ideally one would like to be able to trace formalized claims back to their foundational evidence in the literature just as one does with natural language claims. The micropublications model supports formalization of claims.

Implementation: At the point a formalized claim is created (modeled) from a base statement in the literature, the creating application may capture its supporting statement using the micropublications model. For example, in the current BEL software, instead of capturing only the Pubmed ID of the

publication from which a BEL statement is derived, one might readily capture the backing statement as well, as a micropublication.

Comment: Remember that the minimal form of a micropublication is a simple statement, with its attribution, and the attribution of its encapsulating micropublication.

7. **Modeling annotation and discussion**

Motivation: Annotation and discussion of scientific literature is increasingly conducted on the Web.

Use: Scientific claims and evidence may be annotated in personal or institutional knowledge bases, and may be discussed online in specialized Web portals or communities. Modeling these texts as micropublications, with their backing statements and evidence from the literature, allows them to be exchanged freely between applications in a standard format.

Implementation: This may be done by Web or other applications at the time the publications are annotated or discussion is captured.

8. **Building and using bipolar claim-evidence networks**

Motivation: Scientific discourse often involves disagreement on the correct interpretation or theoretical model for existing evidence. It is important to know where gaps or disagreements exist because these naturally suggest areas for further research. Support/attack relationships exist in the literature for alternative interpretations, hypotheses and models of biomedical function, structure, disease etiology, pathology, agent toxicity, therapeutic action, etc.

Use: We provide an abstract logic representation compatible with much of the current AI literature on argumentation, as well as a description logic presentation modeled in OWL (Additional file 1).

Implementation: Where groups or individuals systematically collect statements and evidence on scientific topics, this may be implemented as a useful pattern.

Comments: We believe this approach could be of particular value in drug discovery and development activities.

9. **Contextualizing micropublications**

Motivation: Micropublications may be applied as annotation to scientific documents, including other micropublications. We use an annotation ontology such as AO [85,86] or OAM [92] to associate micropublication class instances with specific content segments in Web documents, and to record the annotation attribution.

Use: Contextualization is important for the creator of annotations, because it shows them in context. This is of equal importance for the consumer of annotations.

Implementation: Implementation can be within the annotating application.

Comments: We believe micropublications will most commonly be created as semantic annotations on published articles, as this is backward-compatible with the existing publication ecosystem [67].

The most important thing to note about the activities constituting the use cases is that all of them involve assembling, justifying, critiquing, or representing some form or elements of scientific argumentation, including all the support for the argumentation, i.e. including the empirical evidence. Thus, very few of these use cases can be met adequately by purely statement-based models.

Modeling considerations

Constraints on the micropublication model should be imposed not only by the use cases above as they relate to biomedical scientists, but also by other work in the field of argumentation models (as reviewed in [53]), which we would like to be able to reuse where possible. We would like to model both the internal structure of arguments, to digitially summarize publications in a useful way; and the interargumentation structure, so as also to model relations between arguments considered without regard to their internal structure. Also, we want to enable construction of claim networks similar to that described by Greenberg [3], which are principally based on support relationships, but may also have a significant challenge or attack component.

While Toulmin [44] and Verheij's [45,46] approaches might suggest having different relationship types between entities in the model, and the use of *backing* and *warrant* as classes; we avoid that, because these concepts become relativized across a large network: one publication's *backing* is another's *warrant*.

This relativization suggests a graph structure, which is also compatible with work in unipolar [47] and bipolar [49] argumentation frameworks and in claim-evidence networks à la *Hyper* [93] or SWAN [20]. Using a graph model requires common connective properties to allow transitive closure. So for example, the relation between data and its interpretation in a textual statement is called *support*, as is the relationship between a statement and the reference cited to justify it. In another context, these might be modelled as disparate properties, say as "interpretation" and "citation".

Results

The micropublications model is a framework which accomodates a spectrum of complexity, from minimal to maximal representations. It can ingest the simplest forms and give room for stepwise elaboration, consistent with the incremental distributed value chain in which we are trying to embed it. The minimal representation is a single identified statement, where attribution is attached to both the statement and the identification. The maximal representation may be as complex as an entire knowledge base. It is worth stressing that despite the richness of this model, using it does not by any means require deploying all of its concepts for any particular scenario.

To introduce the model, we outline the semantic and mathematical models of argument. Next we illustrate examples of how the model may be applied to each of the Use Cases from the preceding analysis, across the cycle of activities depicted in Figure 2. To illustrate the model, we use exemplar publications as described below. The base classes, predicates and DL-safe rule defintions of the full model, are given in Section A.1, *Class, Predicate and Rule Definitions for Micropublications* of Additional file 2.

Logical formalization of micropublications
Representing arguments

Micropublications represent scientific arguments. The goal of an argument is to induce belief [44]. An argument (therefore a micropublication) argues a principal *claim*, with statements and/or evidence deployed to support it. Its support may also include contrary statements or evidence; and/or the claim may dispute claims made by other arguments. These are called *challenges* in our model, *rebuttal* by Toulmin [44-46], and *attacks* in the artifical intelligence literature on argumentation frameworks (see e.g. [47,49,54,94].

The minimal form of an argument in our model is a statement supported by its attribution. If the source of the statement is trusted, that may be enough to induce belief. Aristotle called this aspect of rhetoric *ethos*, the character and reputation of the speaker [95]. Figure 3 shows this minimal form of micropublication.

The support of a micropublication's claim is structured as a graph. Unlike the standard Toulmin-Verheij model, which only deals with statements, scientific argument must ultimately support statements with empirical evidence, consisting of

- data in the form of tables, images, etc.; and
- descriptions of the reproducible methods by which this data was obtained; which may include drawings, photographs, etc.

Scientific argument must also situate its claim in the context of previous work in the domain, of which it takes account – as additional support, or as error to be challenged and disproven. This context is deployed as paraphrases of other published findings (claims), qualified by a citation of the work from which they were paraphrased. Toulmin calls these paraphrases "warrants" (as
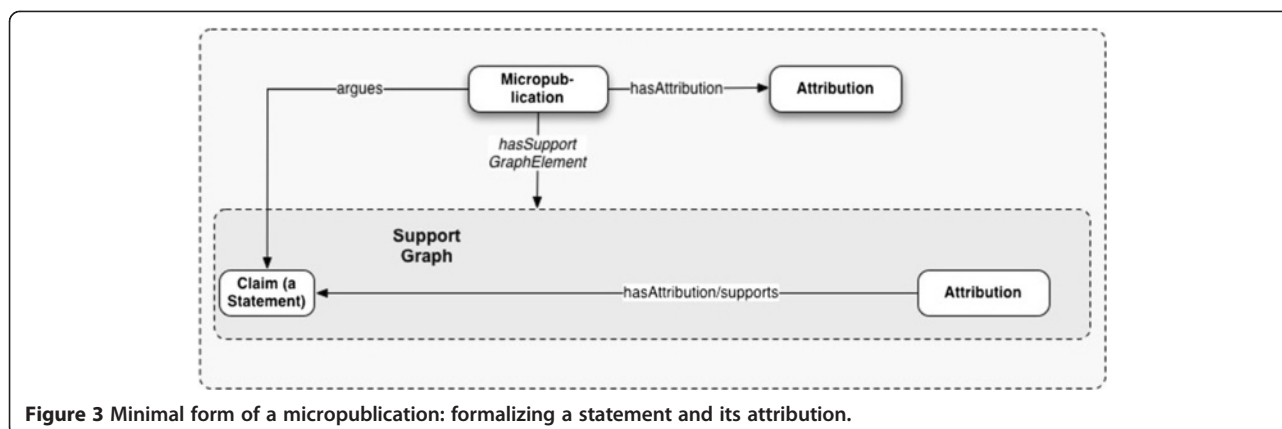
Figure 3 Minimal form of a micropublication: formalizing a statement and its attribution.

in, "warrants for belief"), and the work indicated by a citation is called the "backing", which would be consulted to validate the warrant.

As we are interested in constructing claim networks, it should be clear that in a network, warrant and backing are relative terms. Furthermore, to contruct such a network, we will need to have backing which resides in another work, available in the form of a single statement, not the entire work. While a citation of an entire article may be acceptable as a temporary measure, reflecting pragmatic boundaries, ultimately we wish to have the full claim network at hand. This sets us up to be able to transitively close the network. To do so we use a *supports* relationship between warrant and backing.

Defining this relationship consistently across the model – whether we are dealing with supporting statements, data, or methods – also allows us to bridge the gap between internal argument structure, and inter-argument structure.

We call any element of an argument, a *Representation*[a], a class whose subclasses include *Sentence, Statement, Claim, Data,* and *Method*. *Sentences* need not be syntactically complete – they may consist of a phrase, single word, or single meaningful symbol (e.g. "We hypothesize that", "Often", "¬"). Declarative Sentences are Statements. Sentences which qualify a Statement are *Qualifiers*. The principal Statement in an argument is called a Claim. Statements may be supported by other Statements, or by Data. Data in turn may be supported by Method, i.e. a description of how the Data was obtained, in the form of a re-usable recipe (or recipe component). A *Procedure* or a *Material* is a Method.

Figure 4 shows a simplified model of an argument using this approach, and without deeper semantic characterization of its elements. The Claim is supported here by both a Statement paraphrasing another finding in the literature, and by Data. The paraphrase is supported by a Reference to the work in which we are supposed to be able to find its source. The Data is supported by its Method. Both the micropublication itself, and the

argumentation it formalizes, have Attribution. All *elementsOf* of the Micropublication which support its Claim, are in its *SupportGraph*.

Later we will examine various forms of argument formalized as Micropublications, using a closely related set of examples taken from the literature on Alzheimer Disease, for each Use Case.

Outline semantic representation of the model

The basic outlines of our model are given here more systematically. An OWL model corresponding to these definitions is available at http://purl.org/mp, and detailed definitions are provided in Section A.2 of Additional file 1.

1. Entities, Agents, Artifacts, Activities and Representations.
   a. *Entities* are things which may be discussed, real or imaginary.
   b. An *Agent* is an Entity that makes, modifies, consumes or uses an *Artifact.*
   c. A *Person* or an *Organization* is an Agent.
   d. An *Artifact* is an Entity produced, modified, consumed or used by an *Agent.*
   e. An *Activity* is a process by an Artifact is produced, modified, consumed or used.
   f. A *Representation* is an Artifact which *represents* something.
   g. A Representation may be a *Sentence, Data, Method, Micropublication, Attribution,* or *ArticleText.*
2. Supports and challenges.
   a. The *supports* property is a transitive relation between Representations.
   b. The *challenges* property is inferred when a Representation either *directlyChallenges* another, or *indirectlyChallenges* it by undercutting (*directlyChallenges*) a Representation which *supports* it.
3. Sentences, Statements, Claims and Qualifiers.
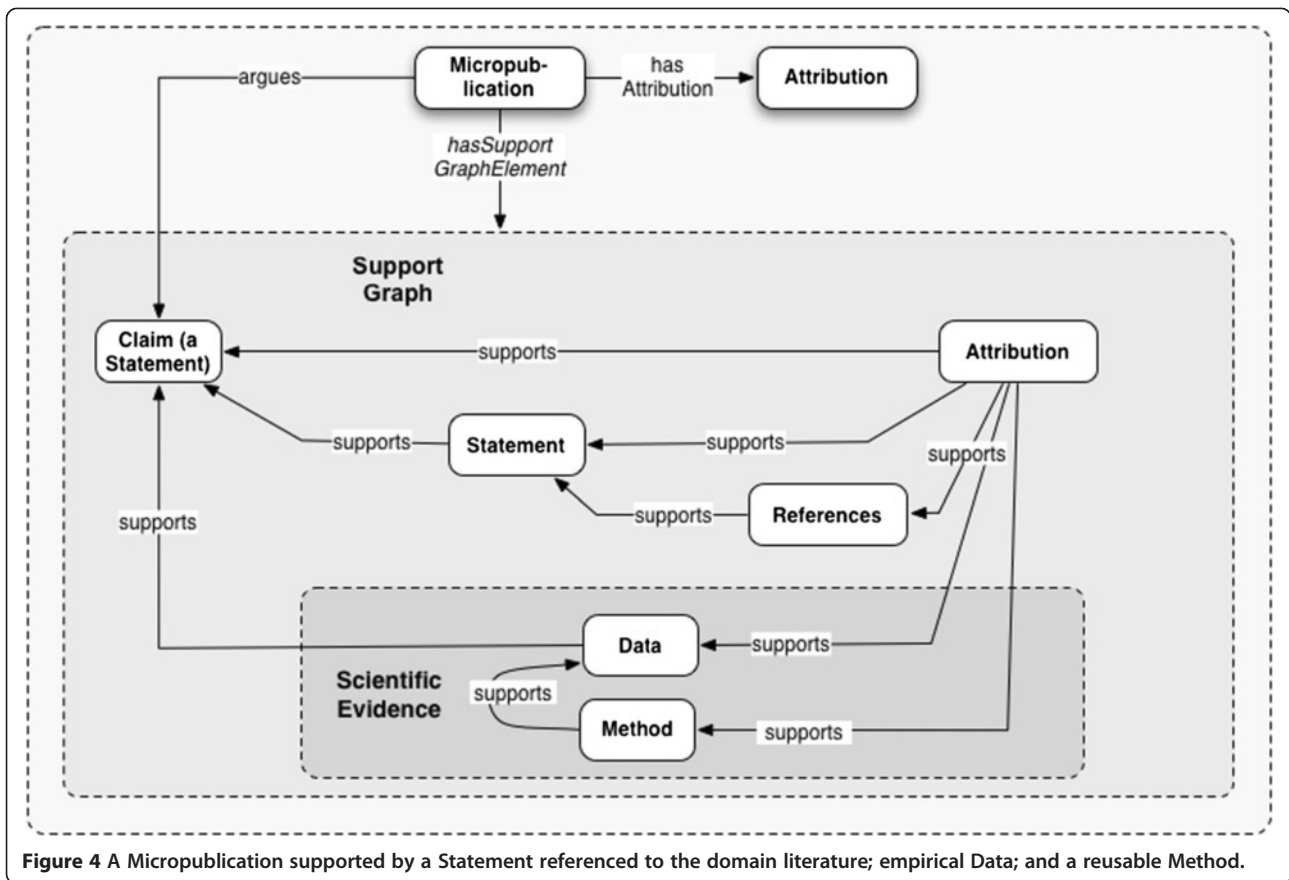
Figure 4 A Micropublication supported by a Statement referenced to the domain literature; empirical Data; and a reusable Method.

a. A *Sentence* is a well-formed series of symbols intended to convey meaning.

b. A *Statement* is a declarative Sentence.

c. A *Claim* is the single principal Statement *arguedBy* a Micropublication.

d. A *Qualifier* is a Sentence, which may modify a Statement. *References* and *SemanticQualifiers* (tags) are two varieties of Qualifier.

4. Data, Method and Material.

  e. Data, Method and Material are kinds of Representation.

  f. If Data *supports* a Statement, that Statement is *supportedByData*. Data may be *supportedByMethod* if a Method *supports* it.

  g. A Method is a reusable recipe showing how the Data were obtained; it specifies an Activity, and may refer to some Material as a component of the recipe. A Material *supports* any Method of which it is a component.

5. Micropublications

  a. A *Micropublication* is a set of Representations, having *supports* and/or *challenges* relations to one another and potentially to those which are an *elementOf* other Micropublications.

  b. A *Representation* is defined as an *elementOf* a Micropublication if that Micropublication either *asserts* or *quotes* it.

    i. A Representation *assertedBy* a Micropublication is originally instantiated by that Micropublication.

    ii. A Representation *quotedBy* a Micropublication is referred to by that Micropublication, after first being instantiated by another Micropublication.

    iii. The asserts and quotes notions are simple extensions of concepts from Carroll et al. and Bizer's work on provenance and trust [96,97].

  c. Claims

    i. A Claim is the principal statement *arguedBy* a Micropublication.

    ii. The *supports* relationships amongst the Representations in a single Micropublications are structured as a directed acyclic graph (DAG), whose root is the Micropublication's Claim.

  d. Attributions

    i. The minimal level of support for any Statement is its Attribution to some Agent.

e. Support and Challenge Graphs
   i. Representations related to the Claim of a Micropublication by the *supports* property, and which are *elementsOf* that Micropublication, constitute its Support Graph, being related to the Micropublication by the property *hasSupportGraphElement*.
   ii. The property *hasChallengeGraphElement* works similarly.

Figure 5 shows the major classes and relationships in the model.

The class Artifact has, as previously noted, a series of subclasses allowing us to deal relatively homogeneously with experimental methods, materials, data, and language artifacts such as statements.

All Statements, as Artifacts, should have an Attribution. The Attribution of a Statement is therefore a part of its SupportGraph. The simplest micropublication would be an instance of the Micropublication class, with its Attribution; arguing a Claim, supported by the Claim's Attribution.

In scientific argumentation, a publication in *Science*, *Nature*, etc. by an eminent scientist, concerning his or her own principal area of expertise, is typically given more weight than an article in a third-tier journal, or a blog post or tweet from an undergraduate with limited expertise. This is why Attribution is part of the SupportGraph of an Argument. However, Attribution alone is weak support. The critical element of support in science is empirical scientific evidence. Complete *Class, Predicate and Rule Definitions for Micropublications* are given in Additional file 2, Section A.1.

Abstract mathematical representation of the model

Let a denote the text of an argument, which in most of our examples will be sections of a scientific article. It may include images and other data as well as text.

Let MP_a denote the corresponding formalization of a as a Micropublication.

Then $MP\_a = \langle A\_mpa, c, A\_c, \Phi, R \rangle$ is the representation of a as a *micropublication* or formalized argument, where

- a ::= an argument text, represented in a document;
- MP_a ::= a micropublication, or formalized argument structure, defined on a;
- A_mpa ::= the Attribution of this formalization of a as a micropublication;
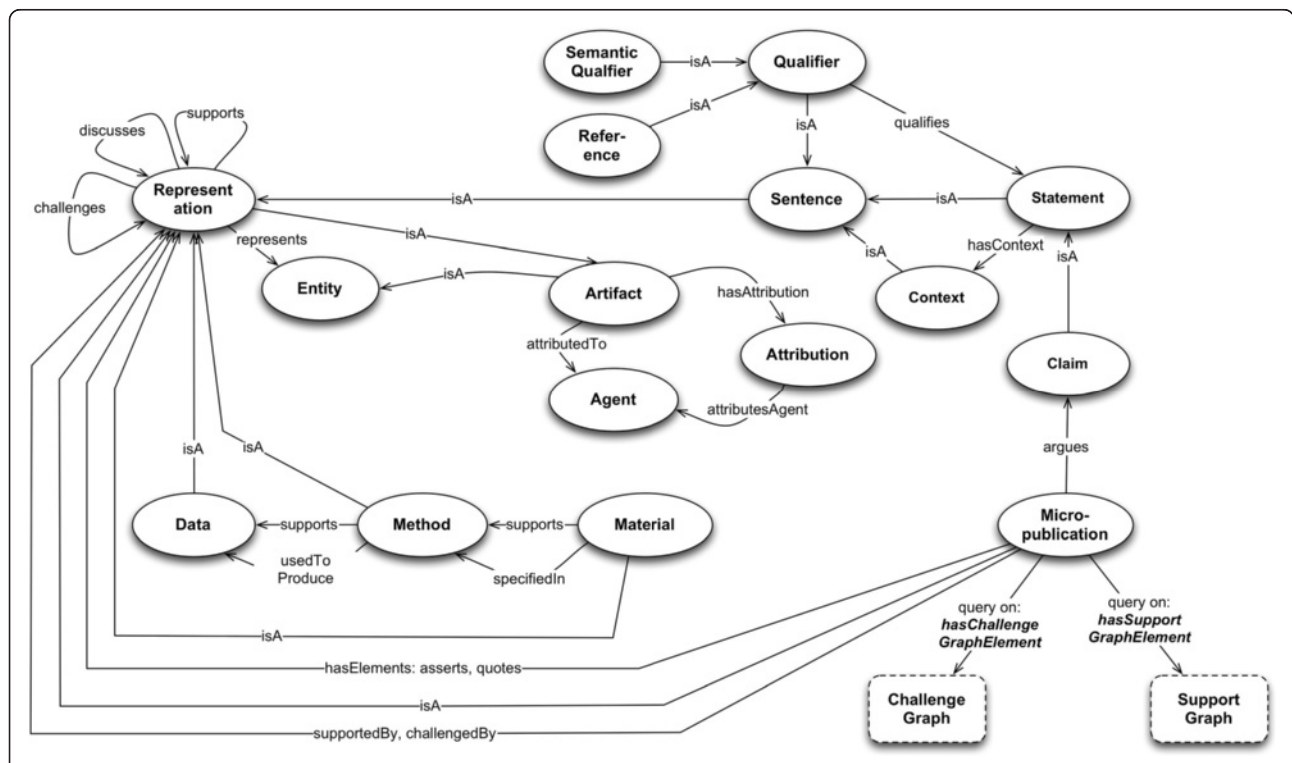
Figure 5 Major classes and relationships in the model. A *Claim* is the main *Statement* argued by a *Micropublication*. A Statement is a truth-bearing *Sentence*, which may be variously *qualifiedBy* some *Qualifier*. A Sentence is a well-formed sequence of symbols, intended to convey meaning; and is not necessarily either complete or truth-bearing. A *Micropublication hasElements* consisting of those *Representations* it *asserts* or *quotes*. A Representation *supports* or *challenges* other Representations. The supporting Representations which are *elementOf* the Micropublication will be in its *SupportGraph*; challenging elements, will be in its *ChallengeGraph*. Dashed-line boundaries indicate graphs instantiated by query.

- c $\in \Phi$ ::= a single Statement, being the principal Claim of *a*;
- A_c $\in \Phi$ ::= the Attribution of the Claim c; and
- Φ is a finite non-empty set of Representations which are elements of the Micropublication;
- R $\subseteq \Phi \times \Phi$; where R is a nonempty disjoint union of supports (R$^+$) and challenges (R$^-$) relations, r($\varphi_i \varphi_j$); $\Phi^+ \subseteq \Phi$ is the nonempty set of all φ_i covered by R$^+$; and R$^+$ is a strict partial order over Φ^+, whose greatest element is c, the principal Claim of the argument in *a*.

Case studies and design patterns

Here we first present an outline of semantic representation of a Micropublication based upon text from Spilman et al. [27] in Figure 6, and then a series of case study examples showing the application of the model to a publication, the structure of its argument, backing in the literature, and concrete scientific evidence. These examples begin with the use case of a single referenced statement and proceed to more complex cases. The models in these case studies constitute a set of design patterns.

Our work builds particularly from experiences in applying the SWAN biomedical discourse ontology [20] to use cases in Alzheimer Disease (AD) research. SWAN led to our development of the Domeo Web annotation toolkit [17,81], and the Annotation Ontology (AO) [85,86] so that formal models of discourse could be readily contextualized directly in ordinary publications, as overlaid metadata. Domeo is currently used to annotate specific reagents in biomedical publications [98], as part of the Neuroscience Information Framework (NIF), among other uses [99]. Other driving problems come from pharmaceutical research, scientific publishing, direct data archiving [100] and data citation [101].

The AD area is perhaps emblematic of the role of competing hypotheses and alternative explanations in biomedical research. There is still not one single accepted explanatory hypothesis, and even what is probably the leading model of AD etiopathology, the "Amyloid Cascade Hypothesis" [102] has seen recent challenges, e.g. [103-106]. AD according to some authors is a syndrome rather than a single disorder [107]. There is still no cure for AD, and simple aging is still the greatest known risk factor. Given this
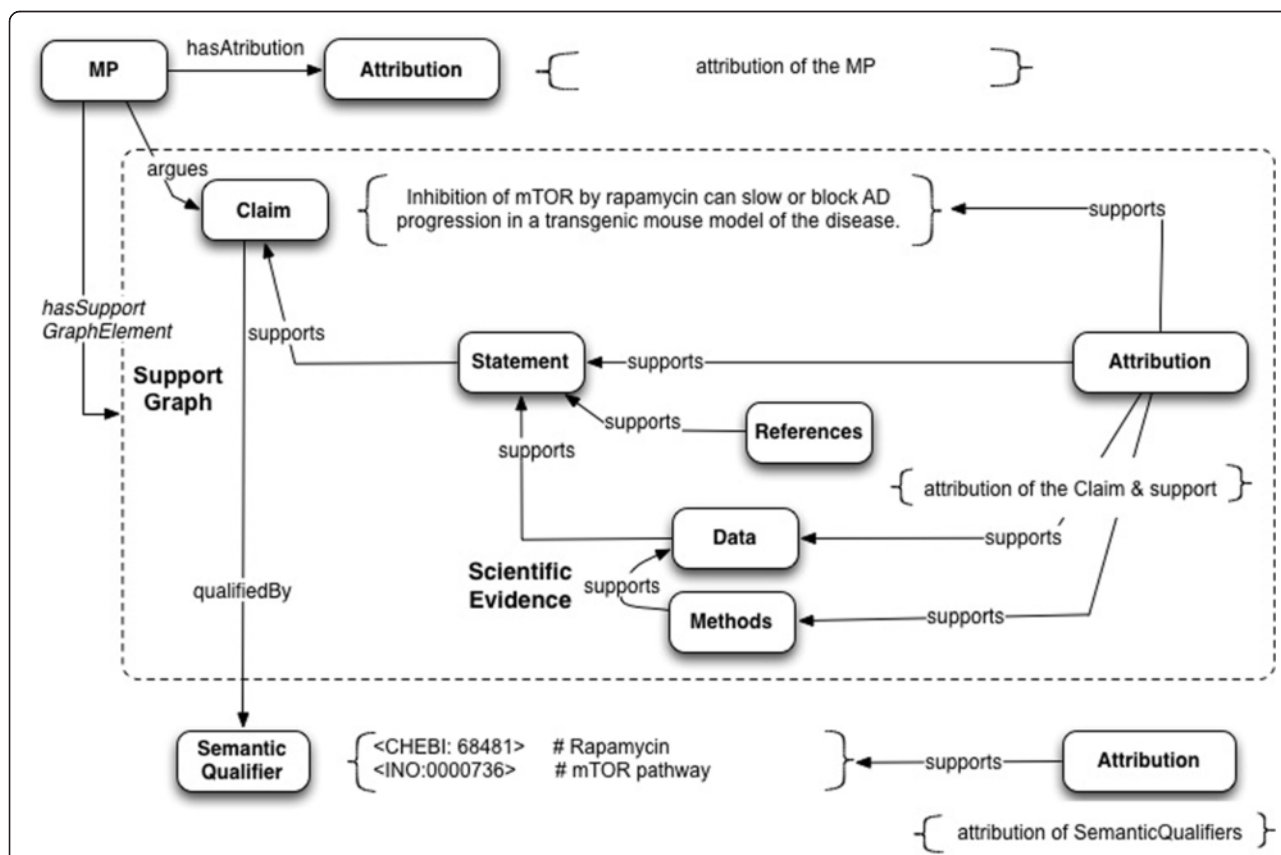
Figure 6 Representation of statements and evidence in a micropublication format, based on text from Spilman et al. [27].
Micropublications show the graph of support for a Statement, including the author's Data and Methods, constituted by a set of *supports* relations in the SupportGraph. The Claim is *qualifiedBy* SemanticQualifiers < CHEBI:68481 > and < INO:0000736>, which are abbreviations for the respective fully-qualified URIs.

background and our prior experience in developing an AD knowledgebase [14,108] and associated ontology of biomedical discourse [19,20] designed to deal with this kind of scientific conflict, we selected our exemplars from AD research.

A recent article by Spilman et al. [27] and its supporting and related materials, are rather typical of scholarly publications from this field. The principal claim in Spilman et al. is that inhibition of the mTOR pathway by rapamycin, in mice genetically engineered as models of AD, reverses AD pathology and symptomatology. This claim is supported in part by references to publications said to demonstrate that rapamycin inhibits mTOR [109]; and that the PDAPP mouse model used in the experiments is indeed a reasonable model of AD [110,111]. The final pieces of support for the claim are (a) data provided by the authors, on rapamycin-fed vs. control mice; and (b) the methods they used to obtain the data, including the feeding protocol [27], the Morris Water Maze protocol [112,113], and the engineered mice themselves [110,111].

We use this and related articles as exemplars in our presentation and discussion.

Figure 6 shows how this semantic representation can be used to model a micropublication based upon text from Spilman et al. [27]. It includes an example instantiation of SemanticQualifiers.

In each Example: below we abstract an argument from the biomedical literature. After illustrating the abstract form it assumes as a Micropublication, we diagram its structure, and describe a use case. In most of the following diagrams, for simplicity, we do not show the Attribution for the Claim and the SupportGraph elements.

RDF examples for several of these use cases are provided in Section A.2.2 of Additional file 1.

Example 1: citable claim with supporting reference and attribution
Use Case: The base use case for Example 1 is constructing directly citable Claims.

Model: The simplest form of a Micropublication represents a Statement with its supporting reference Attribution. In this form it is similar to a nanopublication with text representation of the claim substituted for triples. The argument in this and many subsequent examples is taken from Spilman et al. [27].

Let the argument a_1 = "rapamycin is an inhibitor of the mTOR pathway (Harrison et al., [109])". Then for MP1, the formalized argument (micropublication) derived from a_1, we have

- the Claim (C1) is that "rapamycin is an inhibitor of the mTOR pathway";
- the Attribution (A_C1) of this claim is to the Agent *PSpilman*;

- the Claim is *qualifiedBy* the *SemanticQualifiers* CHEBI: 9168 and INO_0000736;
- the SupportGraph for the claim is the set {supports (Ref5,C1), supports(A_C1, C1) }.

Figure 7 illustrates the structure of Example 1.

A note on Toulmin-Verheij terminology: The Claim in Figure 7 would be a *warrant*, in Toulmin-Verheij terminology, and the Harrison et al. reference is to the *backing* of this *warrant*. The warrant supports belief in the asserted Claim, as either a key supplement to, or in this case, in lieu of, direct evidence (i.e. scientific data) presented in the argument. This micropublication consists only of literature references. We do not assume a cited reference is necessarily valid – it is a representation which may or may not resolve to a real document. If it can be so resolved, we connect it with a further *supports/supportedby* relationship, to some unique identifer of that document. If the document has Micropublication metadata associated with it, we can then potentially further resolve the reference, to a specific Statement within the document.

Use Case Detail: The base use case for Example 1 is constructing *libraries of citable Claims*.

Example 2: modeling evidence support for claims. citable claims with supporting data and reproducible methods
Use Case: The base use case for Example 2 is to enhance *citable Claims* with *supporting Data and reproducible Methods*.

In Figure 8 we abstract and model a Statement from Spilman et al. [27], supported by scientific evidence, i.e., a Representation of *Data*, and the *Methods* (= "materials and methods") by which it was obtained, including pre-observation interventions and observational context.
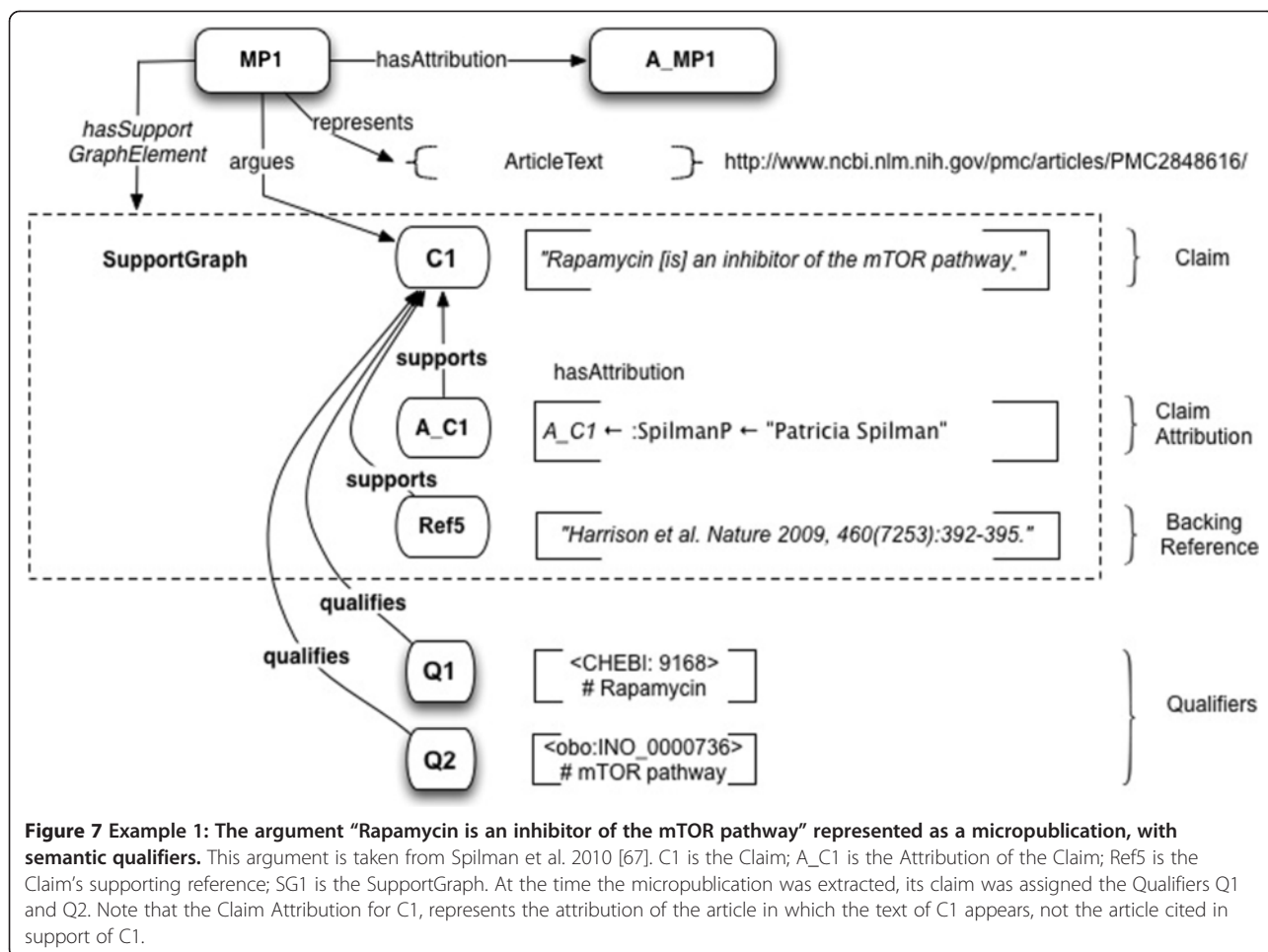
Let a_2 =

- the Claim indicated by C1 in Figure 8 and
- the supporting Evidence indicated by D1, M1 and M2 in Figure 8.

Then for the formalized argument or micropublication MP2, we have

- the Claim is C3;
- the Attribution of this claim is A_C3, which in turn is related to the Agent *PSpilman* by the *attributionOfAgent* property; *PSpilman* is "Patricia Spilman";
- the Support Graph is a set of relations {supports (A_C3,C3), supports(D1,C3), supports(M1, D1), supports(M2, D1)}.

Use Case Details: Citable Claims with supporting data and resources give us the capability of showing the material

Figure 7 Example 1: The argument "Rapamycin is an inhibitor of the mTOR pathway" represented as a micropublication, with semantic qualifiers. This argument is taken from Spilman et al. 2010 [67]. C1 is the Claim; A_C1 is the Attribution of the Claim; Ref5 is the Claim's supporting reference; SG1 is the SupportGraph. At the time the micropublication was extracted, its claim was assigned the Qualifiers Q1 and Q2. Note that the Claim Attribution for C1, represents the attribution of the article in which the text of C1 appears, not the article cited in support of C1.

basis for Claims based on original research in a given work. Also, in the forward-looking case of citable archived data and resources, the claims associated with them might be more easily determined. Citable claims in a personal or institutional "library" or database may "index" the associated data and methods more readily than other forms of metadata. Lastly, methods and materials later found to be flawed might be easily traced to claims based upon them.

Example 3: computable digital summary of a publication

Use Case: The base use case for Example 3 is digital abstracting of a biomedical article in computable form, based on citable Claims with all supporting Statements, Data and Methods.

This more complex form of Micropublication summarizes the principal Claim of an article and the evidence supporting it, again based on [27], and is illustrated in Figure 9.

Let a_3 =

- the Claim indicated by C3 above;

- the supporting Statements indicated by S1, S2 and S3; and
- the supporting Evidence indicated by D1, M1 and M2.

Then for MP3 its micropublication we have

- the Claim is C3, "Inhibition of mTOR by rapamycin can slow or block AD progression in a transgenic mouse model of the disease".;
- A_C3 is the Attribution of C3, with *PSpilman* as its object via *attributionForAgent*, as in the previous example;
- the SupportGraph is a set of *supports* relations on {A_C3, S1, S2, S3, D1, M1, M2, Ref5, Ref9, Ref10}.

Example 4: claim network analysis across publications

Use Case: The base use case for Example 4 is Claim network construction and analysis across publications. This supports the need of researchers to record and see clearly the actual statements and evidence intended to be referenced in a cited publication, as opposed to taking on faith the citation of an entire document treated as a "black box".

Figure 8 Example 2: Scientific Evidence consisting of Data and Methods supporting C3, a Claim from [67]. D1 is a composite image of graphs produced from the primary data. The Methods M1 (a procedure) and M2 (a transgenic mouse strain), both support D1.
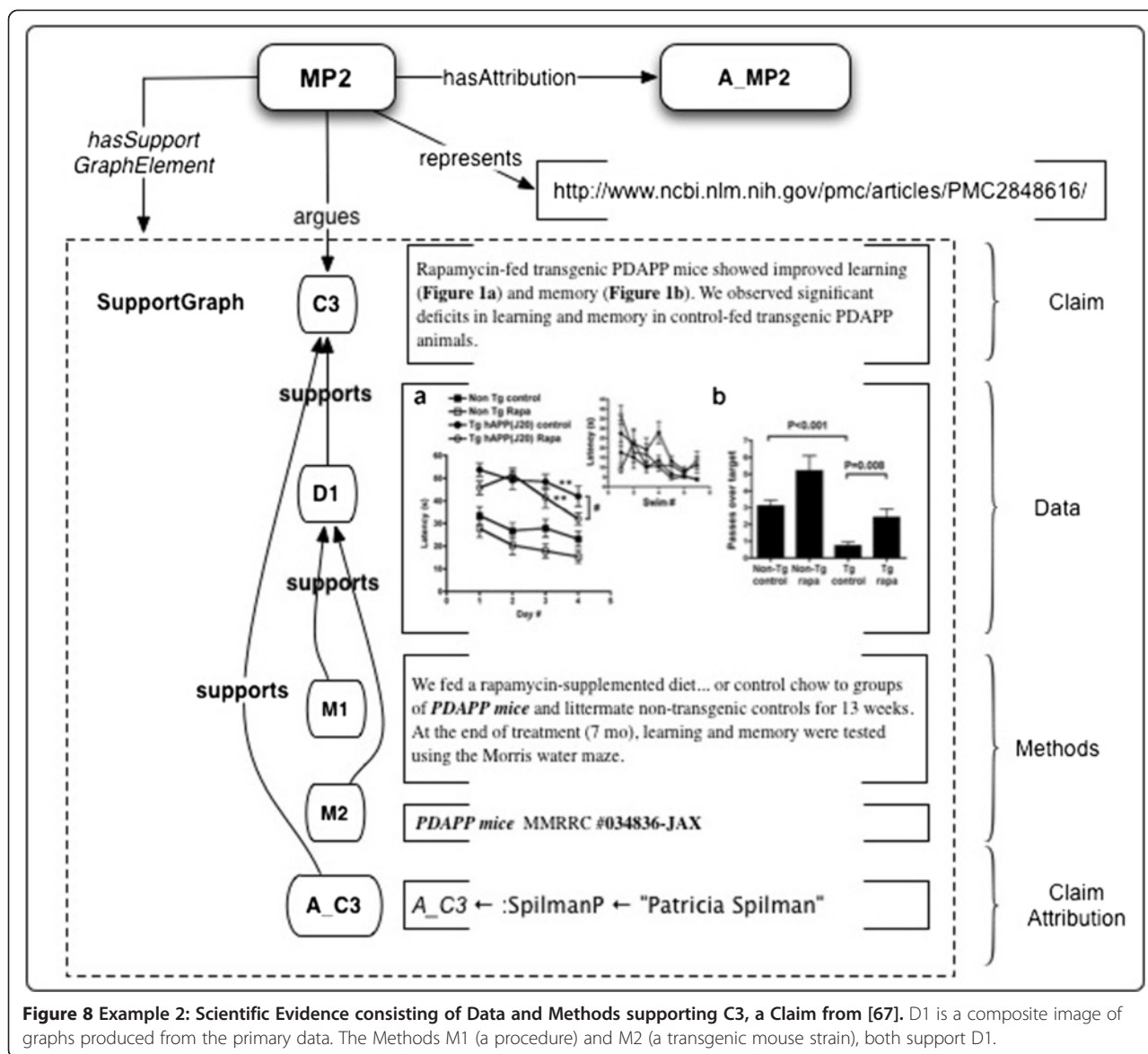
Claim C3 in Figure 10, relies on three logical elements:

1. rapamycin inhibits mTOR (taken from the literature);
2. PDAPP mice are a good (Spilman actually says "established") model of human Alzheimer Disease pathology (also from the literature); and
3. Experimental evidence that PDAPP mice fed rapamycin over time regain some measure of cognitive health.

A flaw in any of these elements undercuts Claim C3, which is the essential argument of the publication. So it is worthwhile to examine, in addition to the experimental evidence, whether the two supporting statements from the literature are well-grounded. Careful readers do this for

important points – or may have the relevant citations already memorized. Implementation of the present use case allows this information from the literature to be modeled, retained and shared.

Suppose that for the supporting literature references in Example 3, [109] and [111], we had also created micropublication annotations over their target documents. It is not unrealistic, given some appropriate software to be used by researchers when looking up the desired claim in a work to be cited, to record it as a citable Claim.

Resolving the Claims in Spilman's argument to their support in the backing references, and connecting the citable Claims, gives the graphs shown in Figures 10 and 11.

The citing Statements in [27] are now connected directly and transparently to their backing in other publications at the level of Claims. The backing Claims in the
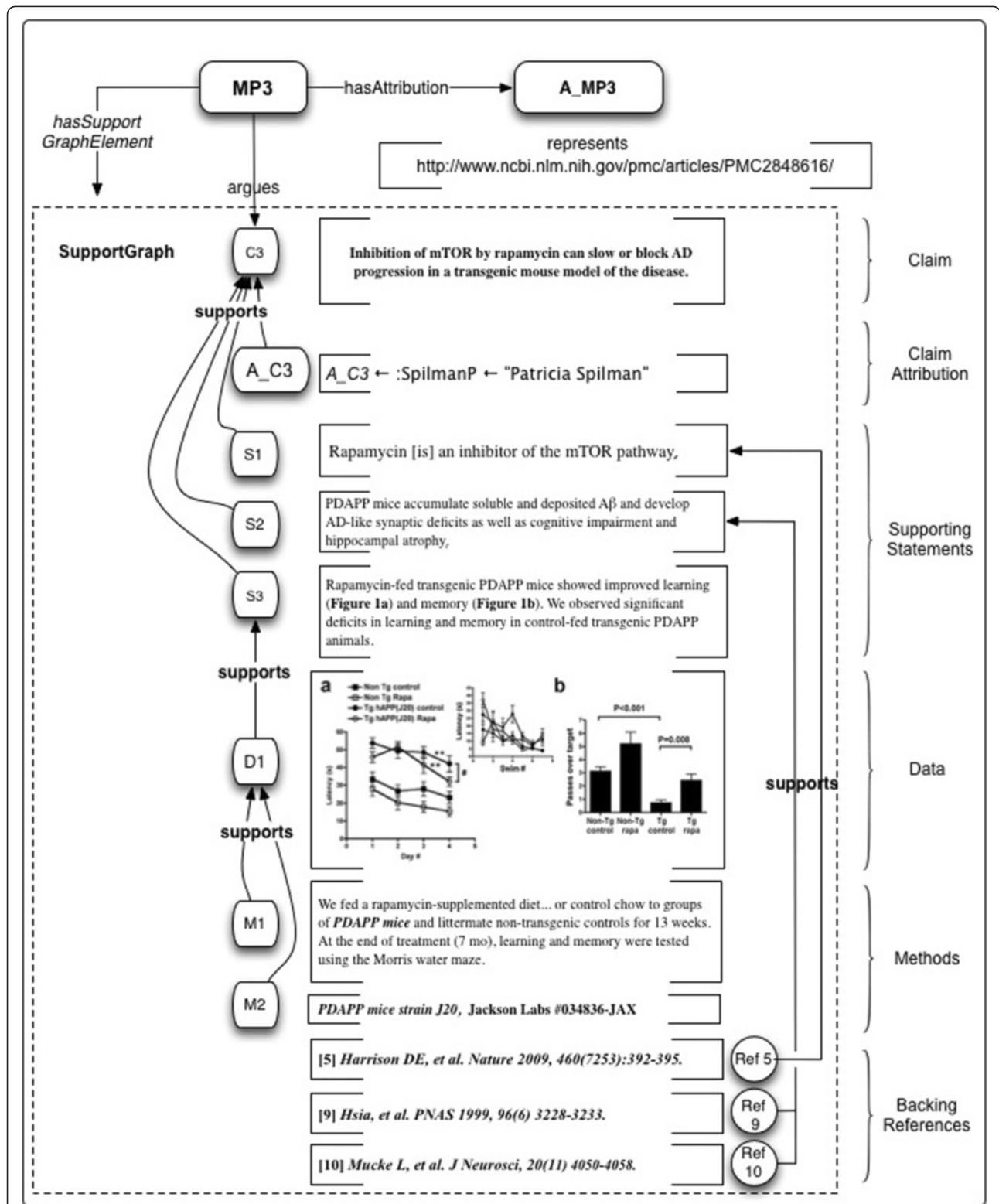
Figure 9 Example 3: Micropublication for the principal Claim of Spilman et al. 2010 [67]. Claim C3 from the abstract is supported by Statements S1 and S2, with support in the literature (Ref5, Ref9 and Ref10); and S3, which interprets the Data in D1. D1 in turn is supported by Methods M1 and M2. All elements of the Support Graph have Attribution A_C3; for simplicity we show only the relationship *supports*(A_C3, C3) in this figure.

Figure 10 Example 4: Transition from document-level backing references, to claim-level backing references, to construct a citation network.
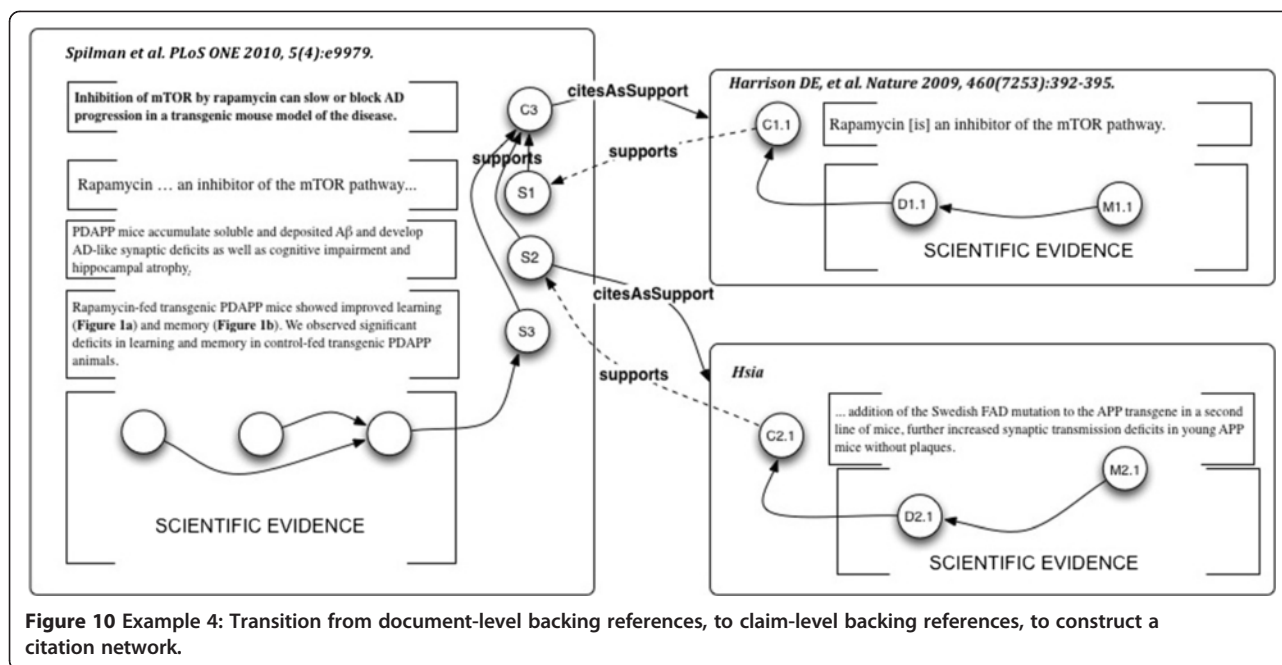
two cited publications are in turn connected directly to supporting primary scientific evidence, which may be inspected.

Spilman's Statement S1 ("rapamycin … an inhibitor of the mTOR pathway…") cites [109] in support; inspection of the cited article shows the corresponding Claim in

Harrison et al. ("rapamycin … an inhibitor of the mTOR pathway…"), shown as C1.1 in Figures 10 and 11.

Spilman's Statement S2 ("PDAPP mice accumulate soluble and deposited Aβ and develop AD-like synaptic deficits as well as cognitive impairment and hippocampal atrophy") cites [111], shown as C2.1 ("…addition of the
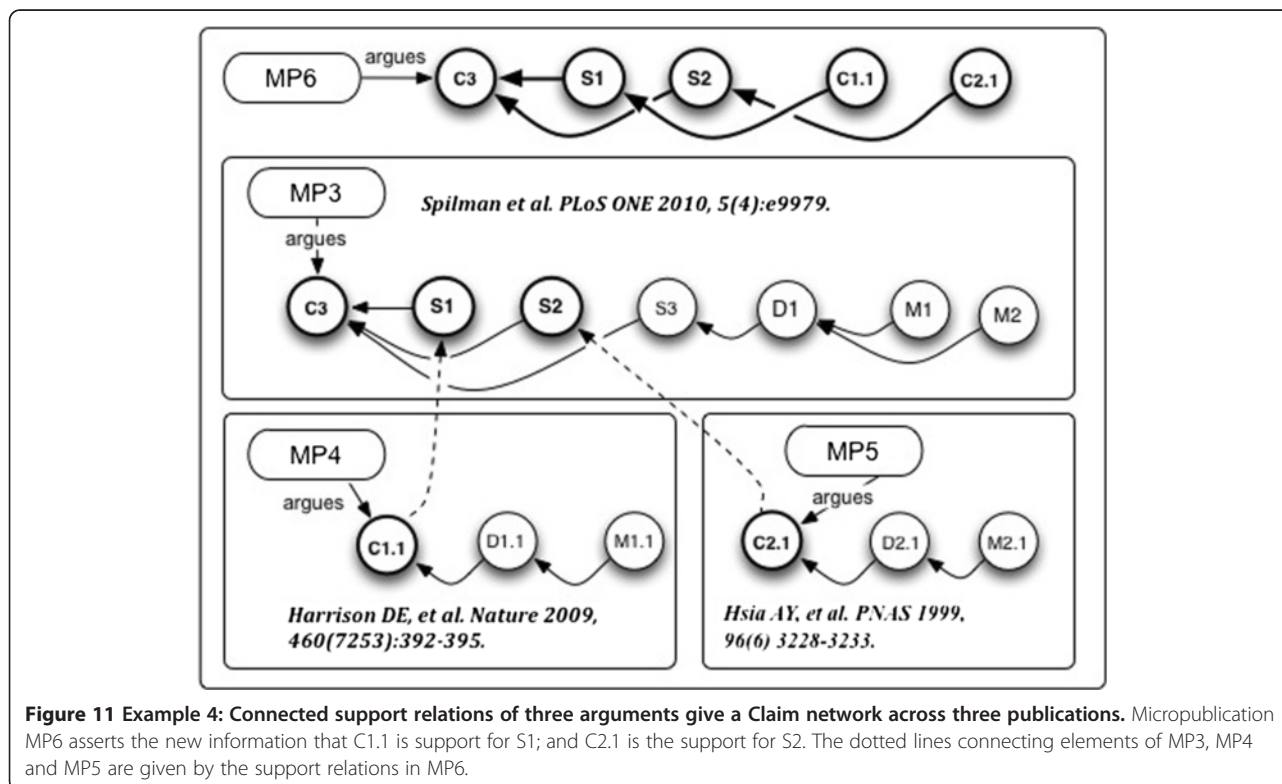
Figure 11 Example 4: Connected support relations of three arguments give a Claim network across three publications. Micropublication MP6 asserts the new information that C1.1 is support for S1; and C2.1 is the support for S2. The dotted lines connecting elements of MP3, MP4 and MP5 are given by the support relations in MP6.

Swedish FAD mutation to the APP transgene in a second line of mice, further increased synaptic transmission deficits in young APP mice…") in Figures 10 and 11.

Close inspection of the Backing for S2 shows how valuable it may be to clarify this at a Claim-to-Claim level; and to cite actual reagent catalog numbers as Methods (see [98] for discussion). The "second line of mice" referred to here turns out to be the J6 line of PDAPP mice.

Both the J6 and the J20 lines are from the same lab. But Spilman et al. used the J20 line, not the J6. The J20 line, Jackson Labs #034836, is barely mentioned in Hsia et al., though both come from the Mucke laboratory. This is perhaps why earlier in the article, the authors simply referred to "PDAPP mice" in general in asserting that they were an "established model" of AD.

Why was a different line used than the one discussed in the actual backing of the reference Hsia et al.? Perhaps it is worth investigating: perhaps the J20 line was readily available, but the J6 line was actually documented. What this points to is that the lines are not sufficiently documented in this article, to the point that authors cite them in a way which seems deliberately obscure. We will ask further questions about these transgenic mouse lines later. But already, the model has helped to show the discrepancy between methods citations, and the actual methods used.

In resolving support across Micropublications, we establish a Claims network. However we must ensure we retain localization of responsibility, for what is originally presented in the merged graphs, and what is "imported" from elsewhere. This is done using the *asserts* and *quotes* predicates. In the example, MP6 *quotes* C3, S1, S2, C1.1 and C2.1. It *asserts* the (new) support relationships between C1.1/S1, and C2.1/S2.

Abstract model of the Claim about rapamycin:
Let $a_4 =$

- Claim C1.1 from Figure 10, and its supporting scientific evidence.

Then for MP4, its micropublication, we have

- the Claim is C1.1, "rapamycin … an inhibitor of the mTOR pathway…", derived from "Harrison et al. [109]";
- the SupportGraph is the set of support relations {supports(D1.1,C1.1), supports(M1.1,D1.1)};

giving us a micropublication model for the rapamycin Claim from [109].
Abstract model of the Claim about PDAPP mice:
Let $a_5 =$

- Claim C2.1 from Figure 10, and its supporting scientific evidence.

Then for MP5 we have

- the Claim is C2.1, "… addition of the Swedish FAD mutation to the APP transgene in a second line of mice, further increased synaptic transmission deficits in young APP mice without plaques"., derived from "Hsia et al. [109]"; and
- the SupportGraph is the set of support relations on {C2.1, D2.1, M2.1};

giving us a micropublication model for the PDAPP mice Claim from [111].
Abstract model of the Spilman et al. Statements, with Backing resolved:
Let $a_6 =$

- Claim C3 from Figure 10; and
- the supporting Statements indicated by S1, S2 and S3.

Then for MP6 we have

- the Claim is C3, "Inhibition of mTOR by rapamycin can slow or block AD progression in a transgenic mouse model of the disease.", from Spilman et al. [27];
- the SupportGraph is expanded and now adds connections to C1.1 and C2.1.

We call the graphs C1.1 → S1 and C2.1 → S2, Claim lineages, by analogy with biological lineages.

Example 5: representing statements with similar or identical meaning: similarity groups and holotypes

It is easy to see that C3 and C1.1 mean the same thing. C3 is derived from C1.1 (in Toulmin's terminology, a "Warrant"). We call these Statements *similogs* of one another. Groups of *similar* statements are equivalence classes, defined as having "sufficient" closeness in meaning to a representative exemplar, or *Holotype* Claim.

A Holotype, or representative of the genus, is selected as a matter of convenience and exemplification, to stand for the common meaning of the Statements in a similarity group, as shown in Figure 12.

In addition to the C1.1 → C1 lineage, we have three other publications, [114-116], containing Claims C4, C5 and C6, with similarity to C1.1 and C3; C4 is selected as representative of the group. It could very legitimately have been cited as support by C3. The Sabatini paper [116], source of Claim C6, is one of the three original articles from 1994 published on this interaction and provides extensive primary evidence. It refers to "RAFT1", a synonym for the preferred protein name mTOR, which superseded it in the nomenclature.

Figure 12 Example 5: Similarity Group with representative (Holotype) Claim. The Holotype is from one of three additional publications, outside the Claim Lineage C1➜ S1.
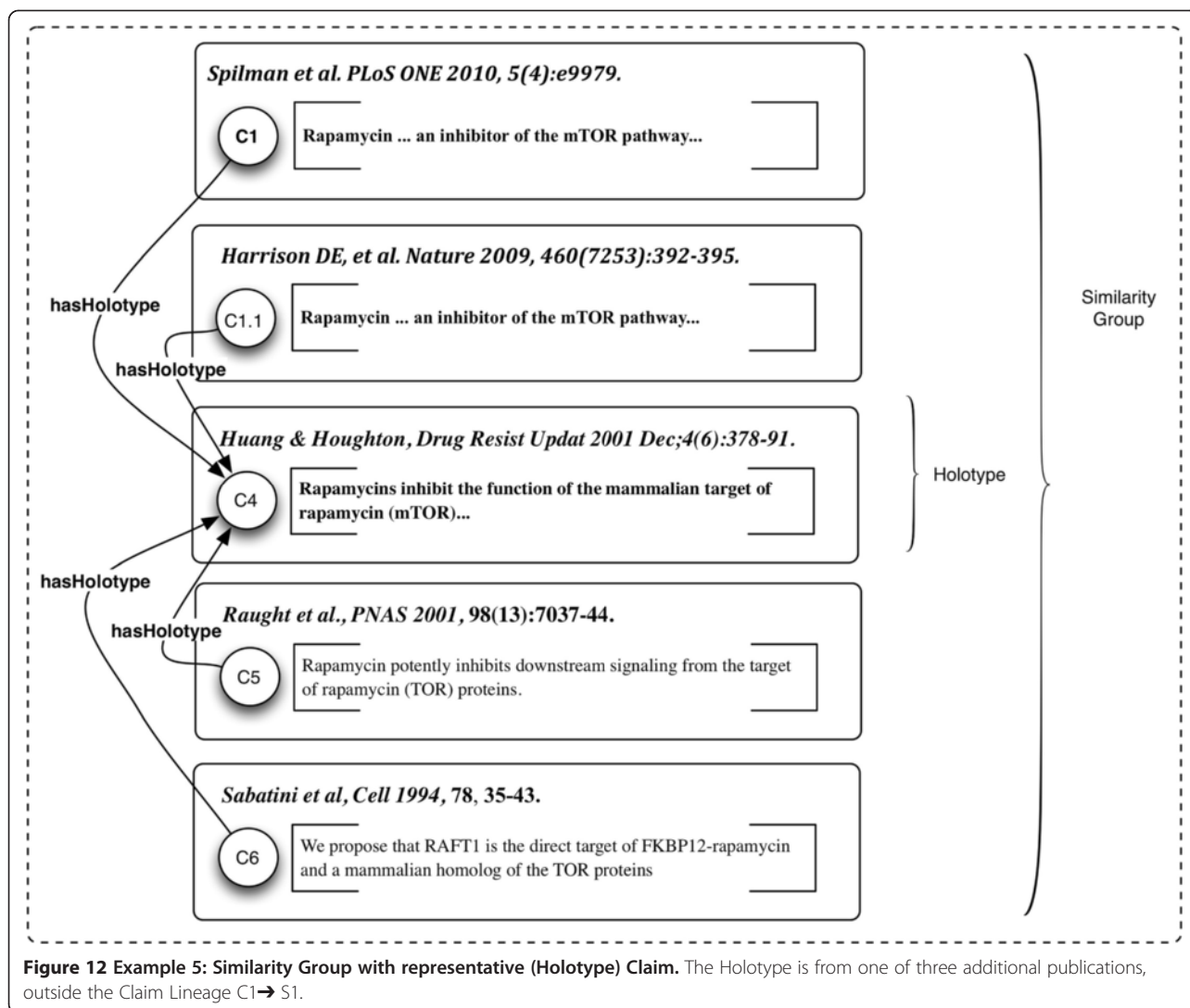
The group-of-similogs/holotype approach is empirical. It is based on the notion that scientific communications in form and content are a "literary technology" [29],which mediates collaboration and exchange of knowledge amongst scientists. As a form of technology, Statements, and the concepts they express, evolve and build upon one another through social interaction. This is a practical alternative to the idea that "Sentences *express* Statements", originating with Strawson's work [117], which considers Statements as extra-mundane ideal abstractions standing apart from the world of real texts as they are exchanged and discussed[b].

Another question now emerges: identification of statements as being *similogs* is itself an assertion. Your "similarity" may not be my "similarity", depending upon both the particular application (think of this as a kind of "manufacturing tolerance"), or in some cases, professional judgement. We can model such an assertion as a micropublication as shown in Figure 13.

In Strawson's perspective, we cannot assign any explicit ontological status, i.e. material existence, to a Statement. It is an extra-mundane abstract "meaning" – but always reduced to "meaning" as expressed in the language of formal logic, which is thus smuggled in via the back door. This would seem to be highly problematic, and we avoid it, because scientific publications for the most part do not represent their findings in this way. We then incur a translation step, which can do violence to the natural language presentation of the actual publication. In our model, Statements are Statements, i.e. declarations, assertions, truth-bearing Sentences, in some meaning-conveying language, and do have explicit ontological status.

Example 6: claim formalization in biological expression language

Use Case: Relating a formal-language model of the content (meaning) of a Claim using a controlled vocabulary, directly to its Backing in the literature. It can be useful
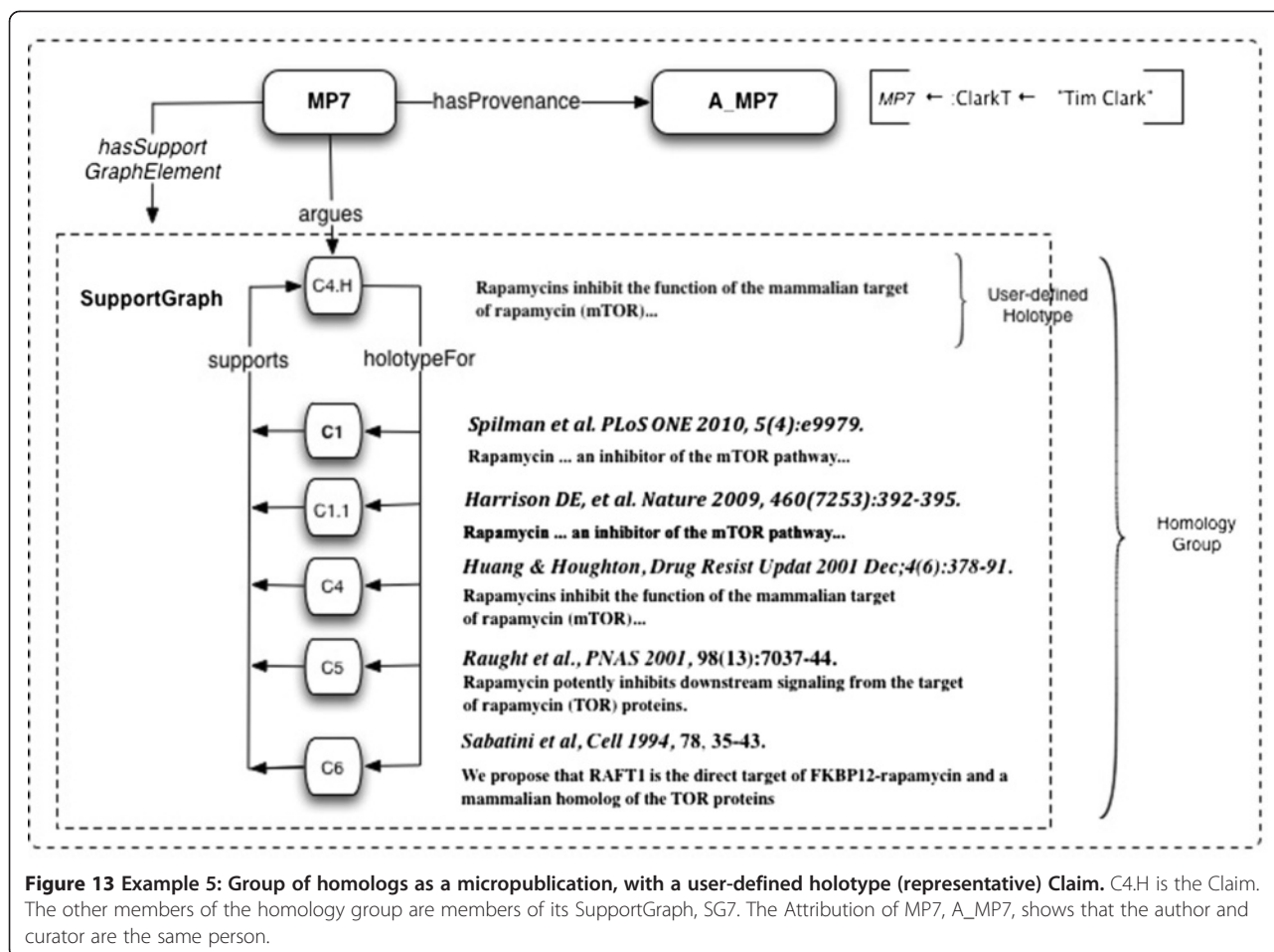
Figure 13 Example 5: Group of homologs as a micropublication, with a user-defined holotype (representative) Claim. C4.H is the Claim. The other members of the homology group are members of its SupportGraph, SG7. The Attribution of MP7, A_MP7, shows that the author and curator are the same person.

to translate the content of textual Claims in a scientific publication, into a specialized formal language, for specific computational tasks - for example, doing systems biology.

The Biological Expression Language (BEL) [18] is one example of such a formalism, which is used to construct knowledgebases of molecular interactions by the pharmaceutical industry. The current software support for BEL associates one or more PubMed identifiers with every BEL statement. As in the case of ordinary textual Claims citing entire documents, it is then laborious to reconstruct the actual Backing when need arises. Yet at the original point of extraction of the BEL statement, that Backing was readily available and could have been cited directly given a suitable model and associated software.

In this case, we model the BEL statement itself as a Micropublication. Here, the Argument Source for the BEL statement is a tuple from a relational database, consisting of the statement text and a reference to the source document, its PubMed ID.

Let a_7 =

- {"a(CHEBI:9168) = | kin(p(HGNC:FRAP1))", "PMID: 12030785"}

We refer to the text of the BEL statement as C7, and to "PMID:12030785" as Ref96, for convenience.

Then for the corresponding micropublication MP7, we have

- the Claim is C7, "(CHEBI: 9168) = | kin(p(HGNC: FRAP1))", which in ordinary English means that sirolimus (also known as rapamycin or CHEBI:9168) downregulates HUGO: FRAP1 (also known by its more common acronym, mTOR).
- A_C7, the Attribution of this Claim, is {"Pratt D, 2013"}, indicating that the Claim was formulated by Dexter Pratt;
- the SupportGraph is the support relation on {A_C7, R96}, indicating the reference to Ref 96 [114] describing the interaction of rapamycin and mTOR in natural language.

As represented in Figure 14, the BEL statement references Ref 96 [114] as support. But this document-level reference can be resolved to a Claim-level reference as shown in Figure 15 and further to a network via its similogs.
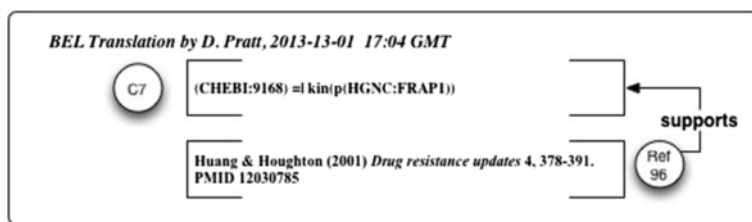
Figure 14 Example 6: BEL representation of a Claim from Spilman et al. 2010 [67] with support at document level only. Attribution of support is based on resolving the PubMed ID from a database.

Claim-level resolution to similogs, would result in a BEL statement for the entire Similarity Group, asserting the rapamycin ↔ mTOR inhibition interaction, whose support thus includes original evidence from the Sabatini group's publication as well as later review information. As the Claim must have been read and interpreted by a researcher to be translated at all, it seems clear that suitable software could embed specific citation to the Claim level at the time the asssociation with the BEL statement is databased.

Example 7: modeling annotation and discussion of scientific statements

Use case: Annotation is a key general use case that associates personal comments, discussion, semantic tags, or other constructs with scientific communication. Micropublications allow Annotations to be associated with logically explicit Backing, while contextualizing

them within the digital content they describe. Readers as well as reviewers and other discussants may – with suitable software support – attach annotations directly to citable claims.

Let's suppose a reader or reviewer of the Spilman et al. article has some notes, comments, or observations to record – for example, about applicability of the Spilman result to drug discovery and development. The reader could then create annotation, modeled as a new, independent micropublication, referencing the original, as shown in Figure 16.

Example 8: modeling challenge and disagreement

Use case: Scientific claims are defeasible. It is critical in evaluating Claims in biomedical research to review critical argumentation which may defeat them. Over time, not only Claims but Methods become established by withstanding such attacks.
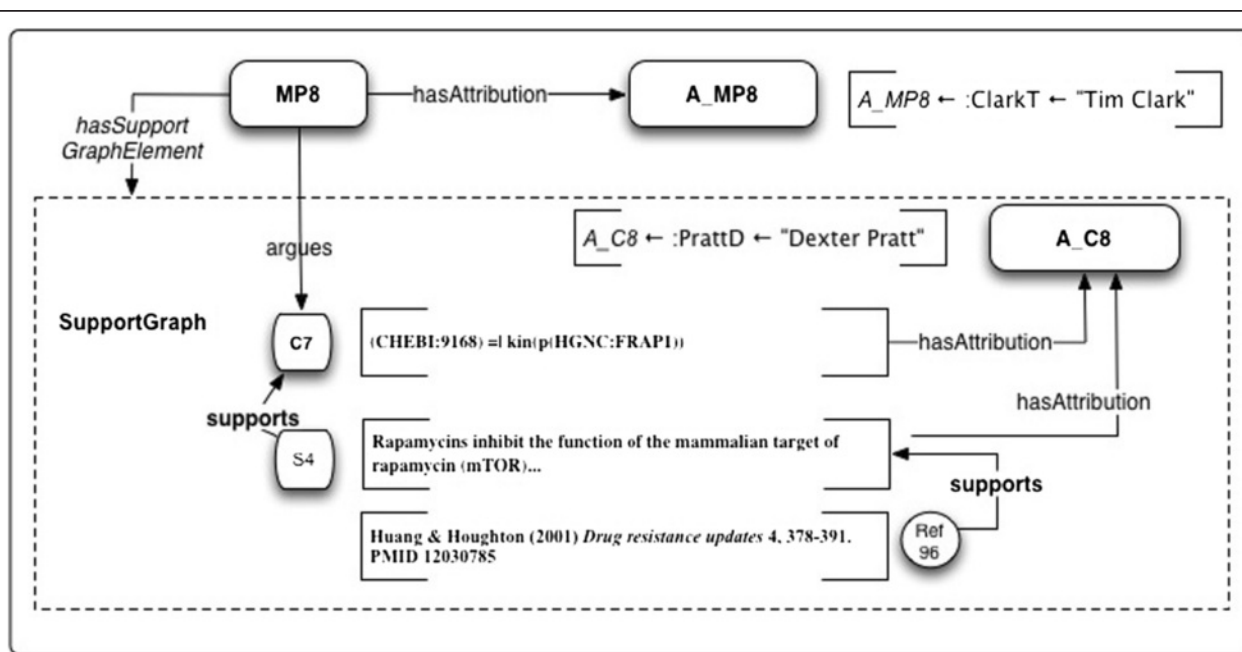
Figure 15 Example 6: BEL representation of a Claim from Spilman et al. 2010 [67] with support resolved to the Claim level. S4 is the specific Claim translated into the BEL statement.

Figure 16 Example 7: Annotation using micropublication relations. C10 is a Claim made by the annotator, "A. Neuroscientist". It is supported by Statement C1 in Spilman et al. 2010, which it annotates.
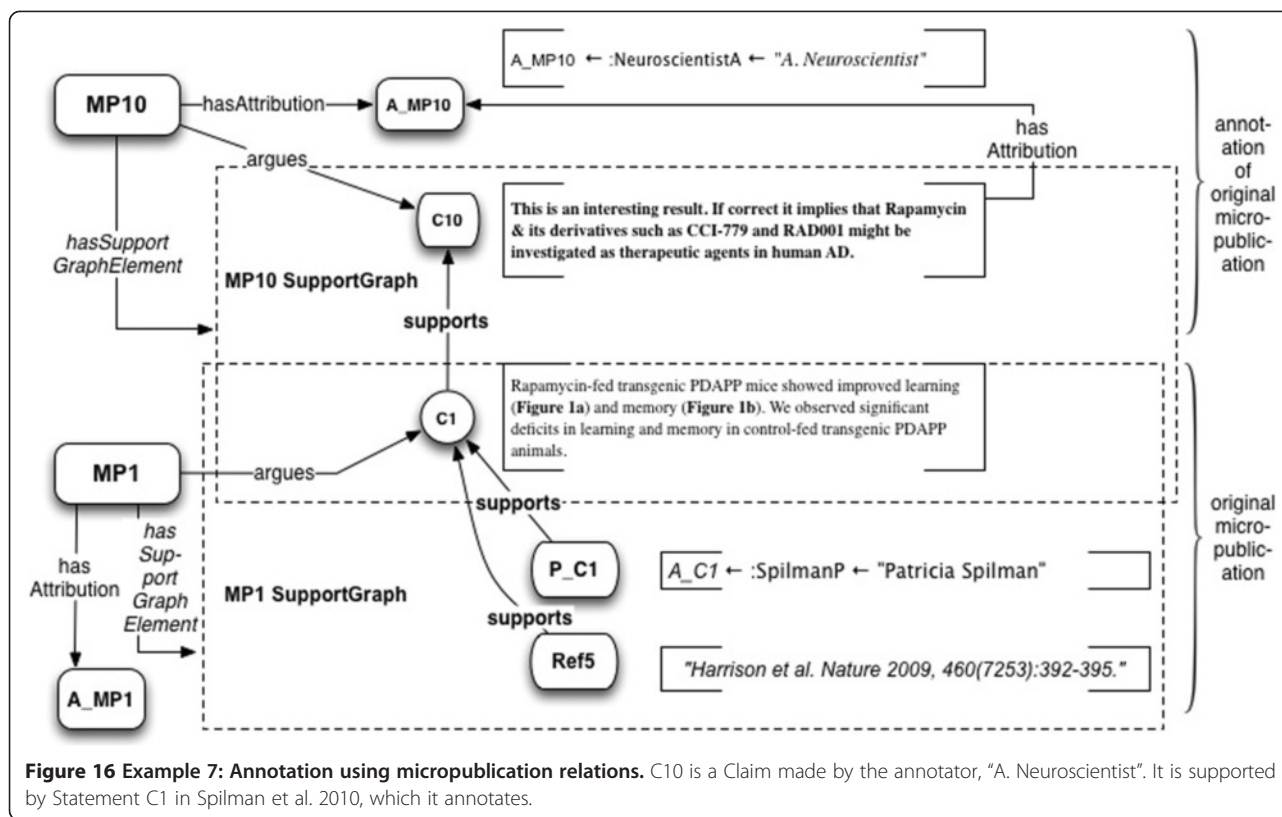
As noted in Example 4, the main Claim of [27] rests on three logical elements, the second of which is

- PDAPP mice are an established model of human Alzheimer Disease.

by which the reader may reasonably infer, that this model is widely accepted as being adequate[c].

Spilman et al.'s principal Claim is still technically valid regardless of whether this is true or not, because the paper wisely restricts it to an assertion about the PDAPP mice only. But if in fact PDAPP mice do not model human Alzheimer's Disease well, the principal Claim as most readers would interpret it, is challenged. The whole point of the article is to suggest that rapamycin or its analogs may be worthy of investigation as therapeutic agents for cases of actual human AD.

Bryan et al. [107], in their review of current transgenic mouse models of AD, make a number of critical points that potentially undercut the methodology used by [27]. One has to do with the low body temperatures of PDAPP transgenic mice. Low body temperatures can cause mice to become hypothermic in the Morris water maze (MWM) protocol; hypothermia impairs performance on the MWM. Since this is an attack on the validity of the model of AD used in the Spilman et al. experiments, in

constructing an extended network of Claims we would like to be alerted to this.

How do we model the challenge of Bryan et al. to the argument in Spilman et al.? In the SWAN project [20], we did so by asserting a formal relationship "inconsistentWith" between Claims. Doing so for any particular pair of Claims, was the task of the knowledge base curator. But the present model does not require a central curator - it is designed to support a collaborative ecosystem, which is a scalable model. So we adopt a different approach, in which annotations of inconsistency between micropublications, are themselves micropublications. This allows consistency/inconsistency relations to be asserted – and selectively consumed - at any point in the ecosystem.

There are two main ways the challenge relationship can be modeled, for two specific use cases. The first way (Case 1) is to model the challenges made within a scientific article, against claims of another publication. The second way (Case 2) is to model disagreement of two articles "from the outside", i.e. from the perspective of a reviewer or annotator. This approach can support as many different views as desired, of the relationships between different publications, and each view can be accorded its own attribution and authorship status, which can be selected for upon retrieval.

Note that *challenges* relationships originate from elements of the SupportGraph of the Micropublication in which they occur. Even challenges initiated elsewhere may be quoted and included, following Toulmin's approach to what he called *rebuttals*.

Case 1: Micropublication models a Claim in one publication that challenges another.

If an author of Publication A explicitly challenges a Claim of Publication B, that is a part of the discourse that can be modeled as part of the micropublication model of Publication A. In such a case, the "external" Claim being argued against will be quoted or summarized within Publication A.

Let a_8 = Bryan et al.'s review. Then for MP8 its micropublication is

- the Claim C11 from [107] as shown in Figure 17, "PDAPP mice tend to have lower body temperatures, which may result in varying degrees of hypothermia during the MWM task, which can produce amnesia in animals".
- A_C11 is the Attribution of C11;
- the SupportGraph is the set of support relations on {C11, R48, R49, R50 };
- the Claim C11 *challenges* statement S3 from MP3.

This would only be a good representation of Bryant et al., however, if the specific challenge to Spilman's paper was actually made in Bryant's text. In this case, it is not. So the observation must be made by a third party. We outline this situation in *Case 2*.
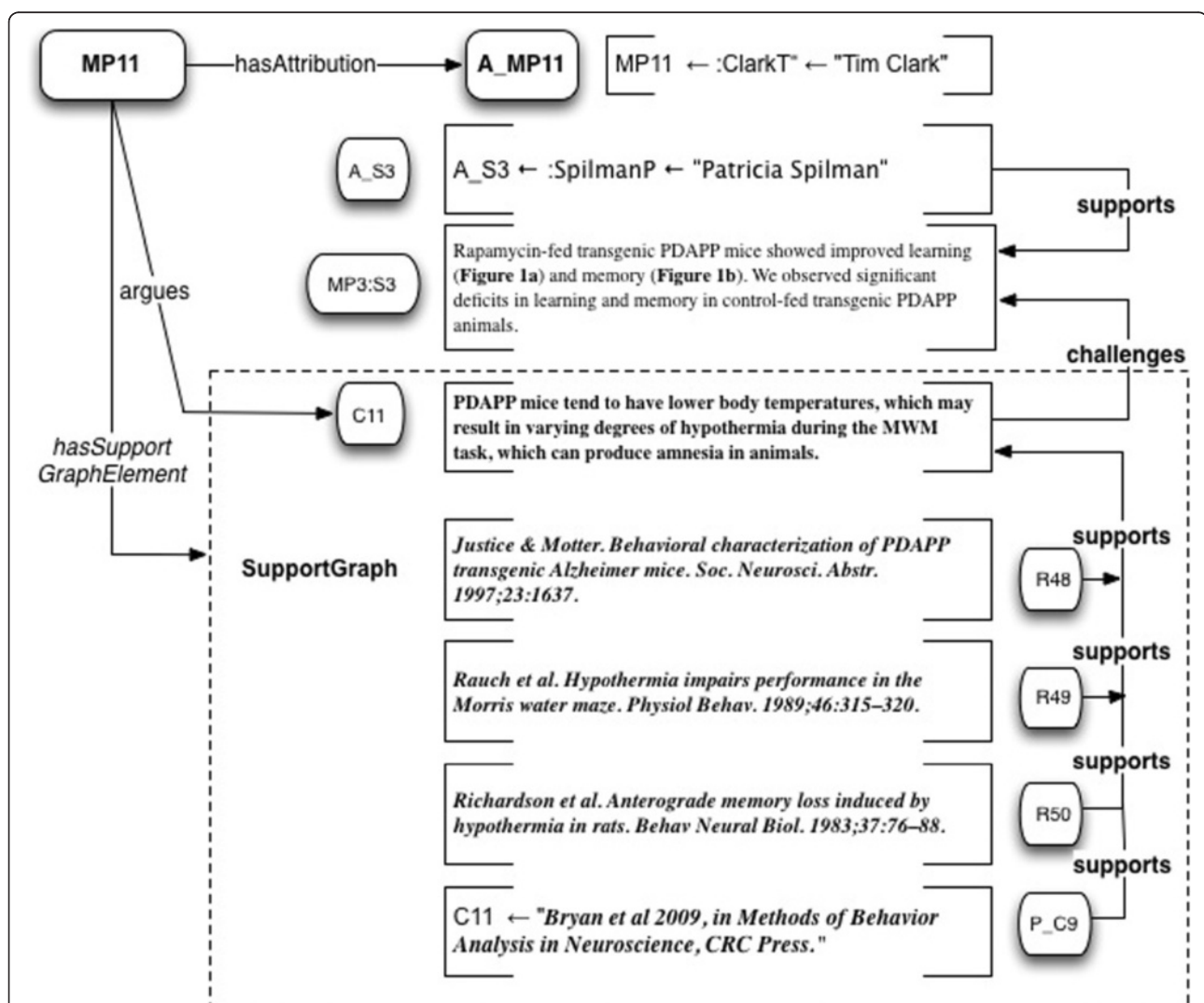
Figure 17 Example 8: Claim C11, taken from [64], Bryan et al.'s review of transgenic mouse models of AD, challenges Statement MP3: S3 in Micropublication MP3. The Claim notes an issue with PDAPP mice tending to confound their performance in the Morris Water Maze (MWM). R48, R49 and R50 are document-level citations.

Case 2: Micropublication models disagreement between two publications "from the outside".

Suppose micropublications MP3 and MP11 are inconsistent, but it is a third party, not the author of either, who takes note of this fact. The initial micropublication summaries do not contain *challenge* relations, but the curator of a knowledge base (KB) wishes to note the discrepancy.

We can create a new micropublication, MP12, as follows.

Let a_{12} =

- A (new) textual assertion that MP11:C11 and MP3: S3 are in conflict; with
- Summaries of the Claims MP11:C11 and MP3:S3.

Then for the MP12 formalization of a_{12} we have

- the new assertion C12, an annotation by "KB Curator", stating: "Bryan et al. claim that PDAPP mice tend to have lower body temperatures, which may result in varying degrees of hypothermia during the MWM task, which in turn can produce amnesia in animals. This challenges the validity of PDAPP mice as an AD model, as asserted in Hsia et al."

This form is diagrammed in Figure 18.

Example 9: contextualization using an annotation ontology

Use cases: (1) Making micropublications visually and experientially a part of the existing communications ecosystem for scientist users. (2) Reliably mapping micropublication components to the documents from which they were extracted, or on which they were expressed as annotations, while (3) allowing them also to exist and be exchanged independently.

Biomedical researchers, and other workers in biomedical communications, spend a lot of time with the domain literature in their field. They read it, write it, discuss it, annotate it, text mine it, give talks and journal clubs about it, and argue about it. Our aim is to enhance and enable these practices, in a way that lets micropublications and other forms of annotation emerge as side-effects from
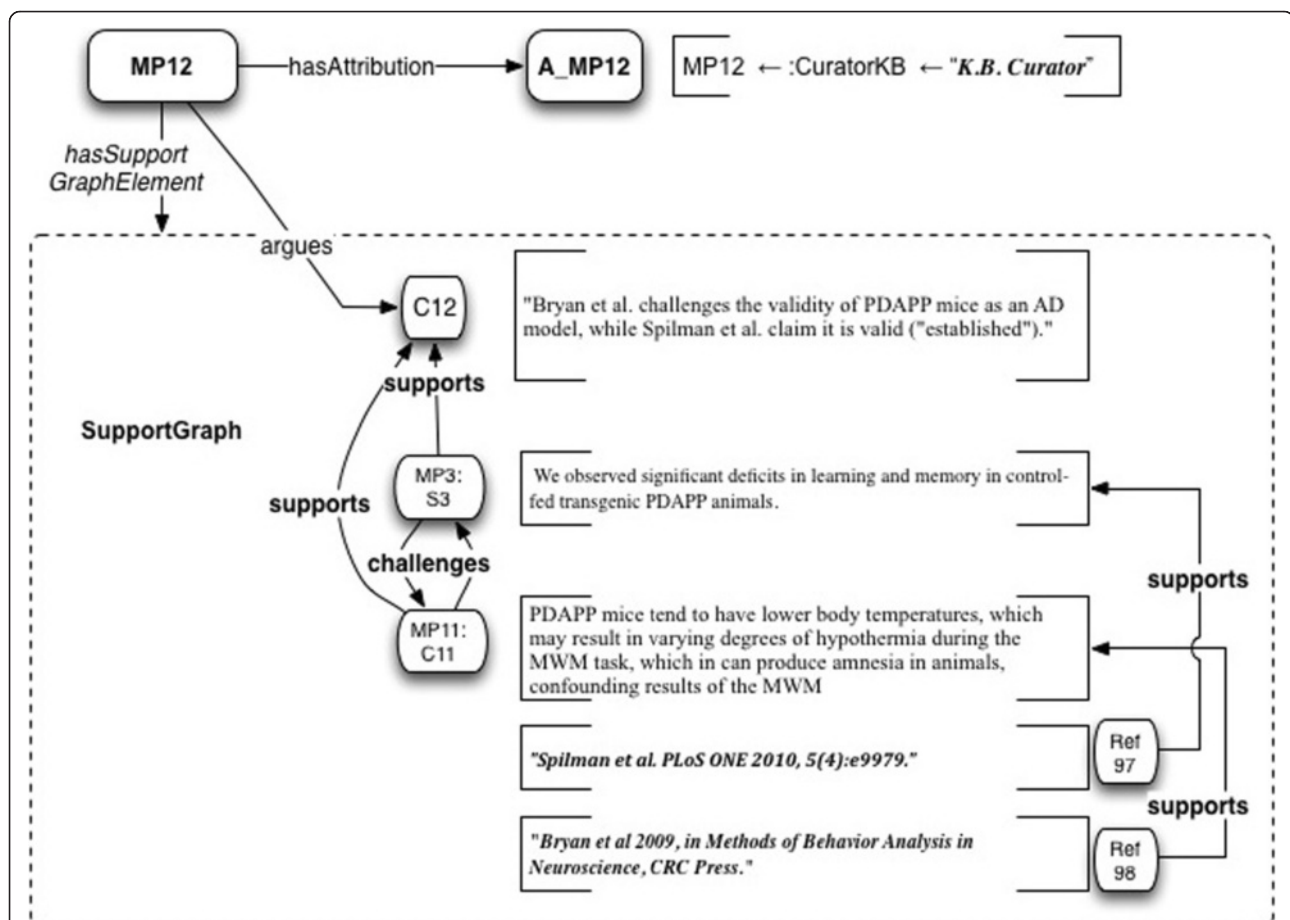
Figure 18 Example 8: Annotation of inconsistency between Spilman et al. [27] and Bryan et al. [107] as an independent micropublication. Challenge relationships between MP3:S3 and MP11:C11 are included in the SupportGraph SG12, because what MP12:C12 argues is precisely that these two Statements are in conflict.

improved practices. This requires (1) the ability to mash up annotations upon the literature they annotate, and (2) the ability to deconstruct annotations by direct reference to segments of existing literature.

We contextualize micropublications using an OWL ontology of annotations, which is orthogonal to the domain ontologies, and to the model of micropublications. Initially we used the Annotation Ontology, AO [85,86]. We have now transitioned to a richer model, OAM (the Open Annotation Model) [87], developed by the W3C Open Annotations Community Group, of which we were founding members[d]. However, the basic principles of these models are roughly the same.

We have demonstrated, using AO, that annotation created on HTML target documents can be exported as RDF, and referenced independently by a PDF viewer (Utopia) [118,119], which successfully mashed it up again into the correct text positions [120].

Figure 19 shows a Claim in a Micropublication contextualized within a full-text article using vocabulary elements from the Open Annotations model.

In addition to the ontologies for annotation contextualization, which can be used by any application, we understand that useful document annotation tools are needed. Our team developed the Domeo web document annotation toolkit [17,81,98], now in version 2 release, for this purpose. Domeo is an open source licensed product (under Apache 2.0) and several groups are now collaborating with us on its further development.

Discussion
Supporting reproducibility in research
Greenberg's publications on citation distortion [3,4] mentioned in the introduction, have great relevance to the problem of reproducibility of results. Begley and Ellis [8] mention two potential major sources of reproducibility problems,

- Cherry picking data, and
- Failure to properly describe methods.

They do not mention citation distortion, of various kinds, which Greenberg showed remains a common and unaddressed problem.

By providing a metadata model that can cite claims directly, micropublications enable citation networks to be constructed using a single common model, which will tend to distribute the costs. This should allow much more clarity to be brought to the question, "on what grounds is this statement made?" Furthermore, the networks may be resolved to purported foundational data and methods, which supply the ultimate evidence supporting a claim.

The elements of a Claim's SupportGraph are its *warrant for belief*, in addition to any empirical supporting
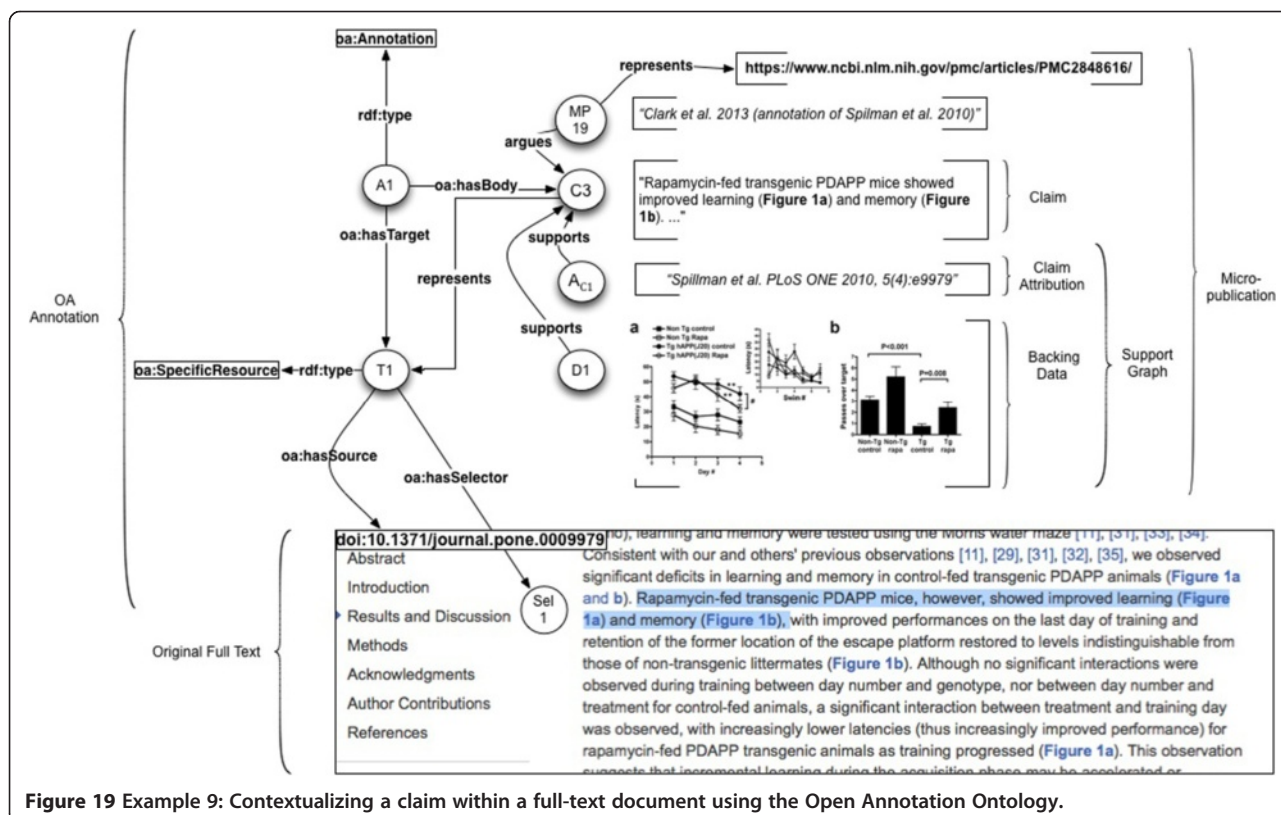
Figure 19 Example 9: Contextualizing a claim within a full-text document using the Open Annotation Ontology.

evidence directly presented. In Toulmin's terminology, the *warrant* is a purported summary of the *backing*. In Example 1 this would be the actual document referred to by the citation [109]. If the warrant does not adequately represent, distorts, or fabricates its backing, and this is known, the validity of the Claim it supports is called into question.

Greenberg demonstrated that this happens far more frequently than one would like and is a serious defect in the domain literature which he reviewed (amounting to hundreds of papers). Because it is far too laborious to check each and every cited document, searching for the relevant Claim, citable Claims are proposed here as a method to dramatically reduce the labor cost of checking a Claim's support. With proper software, identifying such Claims may be automated.

Citable Claims could be constructed most economically at the point where researchers read and take notes upon, or search for backing for their own assertions in, the domain literature of their field. Today researchers commonly use bibliographic reference managers such as Zotero, EndNote, etc. to record this aspect of reading

and note-taking, but the Claim(s) for which they record a bibliographic reference are not captured. Our model allows them to have a standard sharable representation.

Figure 20 shows an example of citation distortion based on a section of Greenberg's supplemental data, represented as a micropublication citation network drawn from eight publications [121-128].

It can be readily seen that Needham et al.'s [121] claim that "[t]he accumulation of APP and its fragments is often stated to precede other abnormalities in IBM muscle fibers…", made in a respected journal, is based only on citations to one laboratory, which repeatedly self-cites, and whose supposed foundational references are nothing but hypotheses, if that. In fact, these authors (excluding the third) wrote another review [129] published that same year in *Lancet Neurology*, which removed the "is often stated to" qualifier, upgrading the claim to "fact", without citation.

Greenberg also notes, in reviewing one of the "foundational" publications from the same laboratory [124], "major technical weaknesses" including "lack of quantitative data" and "lack of specificity of reagents". Both of
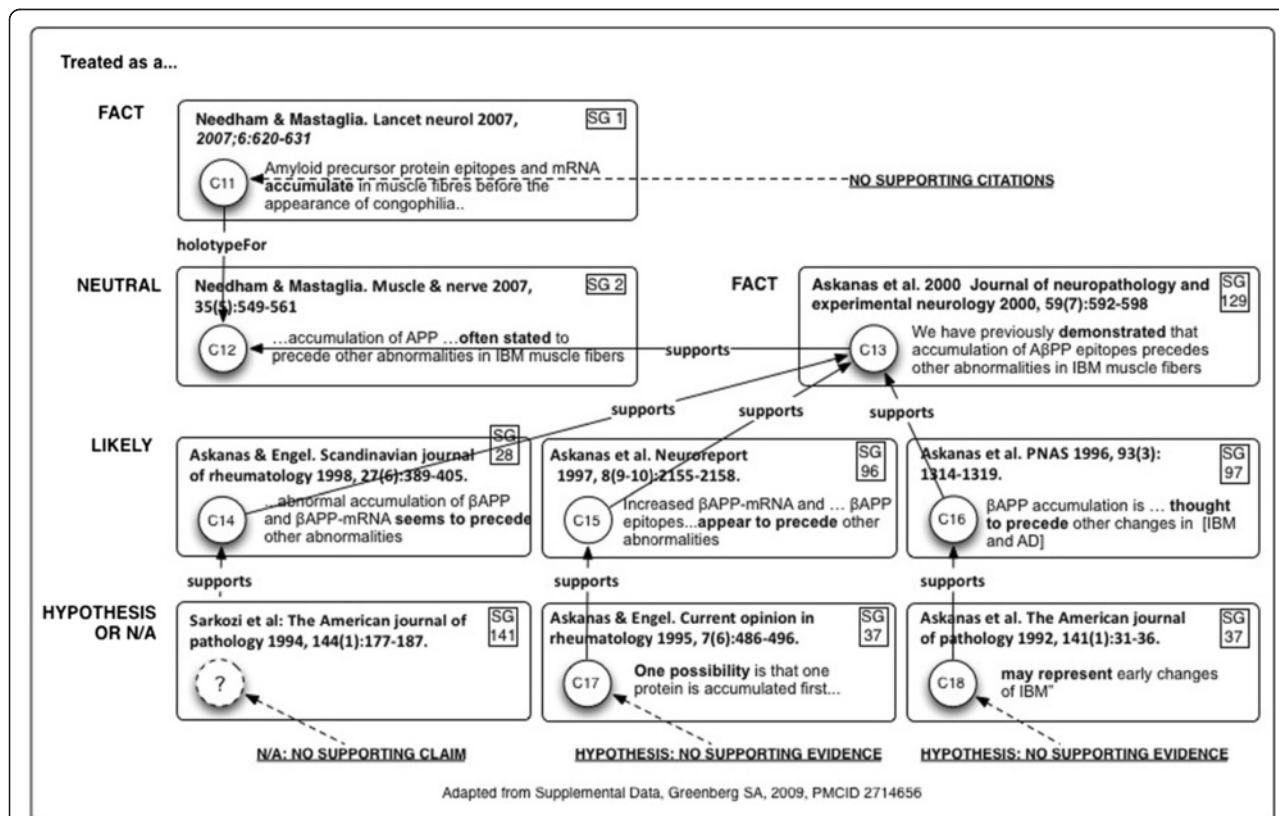
Figure 20 Citation distortion: Graph of a claim network from Greenberg 2009 [3] showing citation distortion. The claim presented in this lineage states that amyloid beta deposition in IBM muscle fibers precedes other pathological changes – from which can be inferred that it is the causative factor. The foundational publications, which would be expected to contain supporting data, do not. Needham and Massaglia's article from Lancet Neurology provides no support at all for Claim C11 at all, treating it as a pure fact. In the Claim network, we assert C11 as a Holotype, or representative Statement, for C12. In fact, that assertion could be made for all Claims in the network.

these problems, particularly that of poor reagent specificity, can be highlighted readily in micropublications, because of the ability to cite both data and methods (reagents).

It is true that reagent citation is a more complex problem than simply providing and implementing this model can resolve. It needs urgent action by publishers to require more specificity in textual identification by authors, along with availability of global registries such as those provided by the Neuroscience Information Framework [98,99,130], for a reagent-citing approach to bear fruit. Our contribution here is to provide an integrated approach to linking reagents and data directly and lucidly to the scientific claims they support.

Resolution of claims to their support in data and methods, including research reagents, is an important new feature, not provided by our previous SWAN model (see Section A.3 in Additional file 3 for a detailed comparison of micropublications to the SWAN model).

Implementation in software

A natural question that arises is, can we embed this model usefully in software? This is required if it is to be effective. A related project in our group, Domeo [17,81,98], allows us to do so. Domeo is a web tookit for automated, semi-automated and manual annotation of web documents. It consists of a knowledgebase, parsers, web services, proxy server and browser-based interface. It has a plug-in based architecture and supports profile-based selection of user interfaces. Domeo is designed to support multiple knowledge-base instances communicating peer-to-peer, and allows annotations to be kept private, group-specific, or public.

Domeo is in active use as part of the Neuroscience Information Framework, is installed for use by drug-hunting teams at a major pharmaceutical company, and by an NIH-funded project developing a drug interactions database. It is open source software licensed under Apache 2.0. It has a growing network of contributors, from both academia and industry.

Domeo version 1 allowed users to annotate web documents using the SWAN ontology of scientific discourse. As noted previously, one of the motivations for developing the Micropublications model was to address short-comings in SWAN by means of a systematic re-thinking of the model, based on a different starting point.

We have now also implemented a Micropublications-based annotation functionality as a plugin to Domeo version 2. The user interface (UI) for this plugin is shown in Figure 21, using the Spilman et al. [27] example again. A user begins by defining individual Statements and their support. This is done by highlighting the desired Statement, clicking "annotate" and selecting "micropublications" as the type of annotation. The Statement then

appears in the panel and the user is given a choice of connecting it via the *supportedBy* relation to (a) its References, which have already been parsed out as computable objects; (b) Data, in the form of images in the article, also already parsed out as computable and annotatable objects; or (c) other micropublication-annotated Statements within the article.

Currently Domeo implements an internal version of the micropublications model which it can serialize and store persistently; this version is currently deployed.

One of the key features the micropublications model will provide in the context of pharmaceutical research, is the ability to securely link proprietary internal data as supporting evidence.
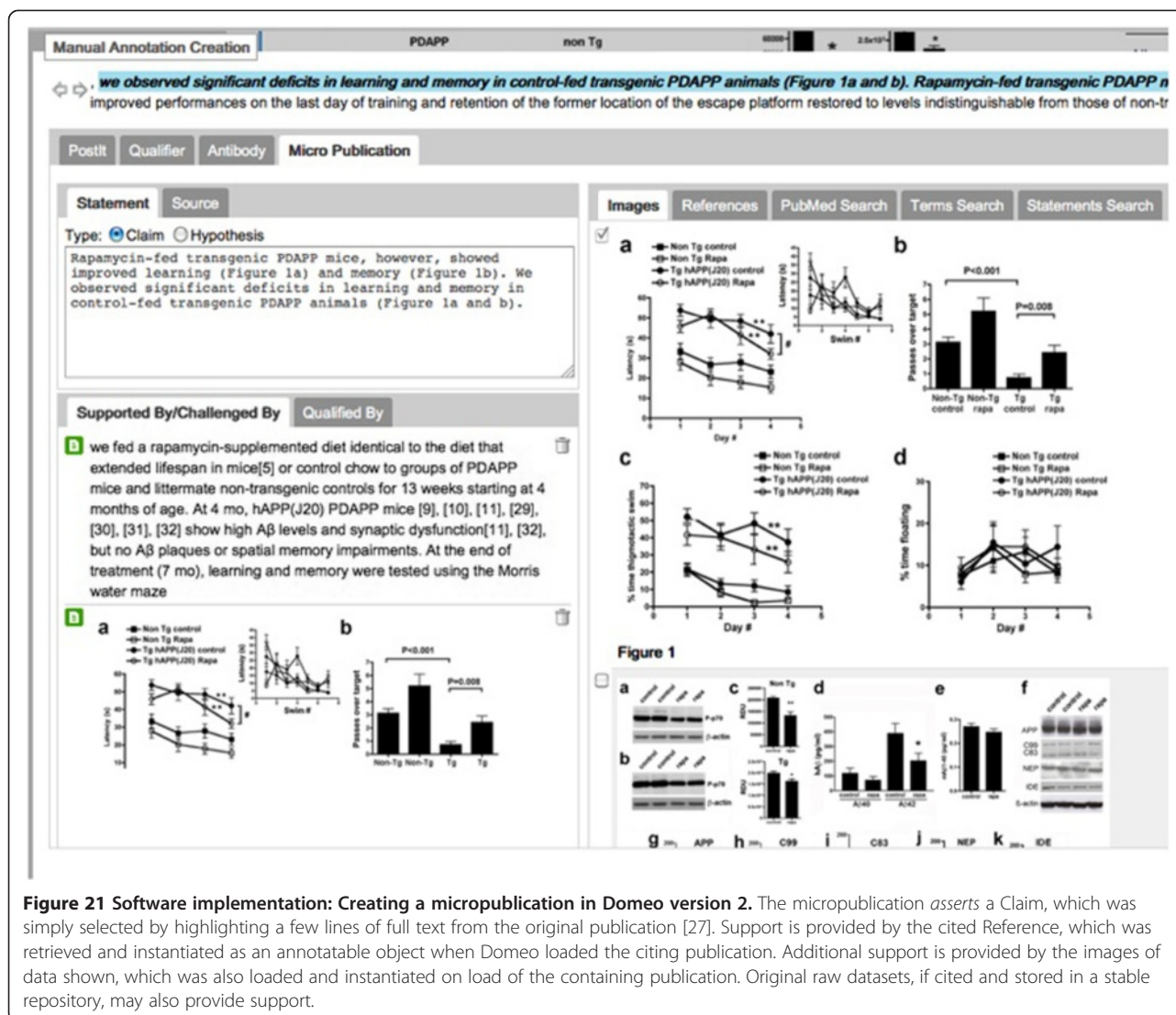
We proposed this model as a general purpose representation model for claims, evidence and argument in the biomedical and other scientific literature. By design it is not intended to be limited in applicability to a single program. To be most useful, it requires reasonably widespread adoption. In addition to the Domeo implementation we describe above, we suggest two particularly attractive points of uptake to the community at large.

1. Any bibliographic management software can implement micropublications. This would enable the scientist user to have a structured library at hand, not only of the references s/he has cited, but also of the particular statements within the referenced material, which are most important.
2. Publishers may implement micropublications as value-added annotation to their published content.

The main value proposition for implementing this model is to enable individual statements in the literature to be cited and referenced directly both by traditional articles and in web discussions,; linked easily into citation networks; and grounded in their supporting evidence – data and methods. We believe this is a pressing need across the biomedical research and development community, to provide better accountability and to support a culture of greater transparency. A number of poor scholarly practices regarding evidence might also be prevented by widespread adoption of our model, embedded in useful applications. Ultimately we believe it can greatly facilitate data and methods re-usability, improving the reliability and reproducibility of research results.

Conclusions

Micropublications enable us to formalize the arguments and evidence in scientific publications. We have shown the need for this kind of formalization in order to meet a number of significant use cases in the biomedical communications ecosystem. We have also demonstrated that purely statement-based models are too underfeatured to

Figure 21 Software implementation: Creating a micropublication in Domeo version 2. The micropublication *asserts* a Claim, which was simply selected by highlighting a few lines of full text from the original publication [27]. Support is provided by the cited Reference, which was retrieved and instantiated as an annotatable object when Domeo loaded the citing publication. Additional support is provided by the images of data shown, which was also loaded and instantiated on load of the containing publication. Original raw datasets, if cited and stored in a stable repository, may also provide support.

deal with scientific controversy and evidentiary requirements, and therefore cannot adequately meet use cases requiring an examination of evidence – which we believe will be the majority of those arising from the primary literature.

Our analysis clearly places the Micropublications model, or some near relative, among the required components for next-generation scientific publishing.

Nonetheless we believe statement-based models also have an important role in formalizing statements, or at least those which have already been deployed with an adequate set of supporting evidence. Such formalizations should also record the original authors' textual interpretation of the empirical evidence as the grounds for later formalization. We provide a clear interface to such models.

An example of this kind of application would be in drug hunting teams in a pharma or biotech company. Examination of evidence and possible contradiction is absolutely fundamental to the activities of such teams in qualifying targets and leads. At the same time, with adequate evidence chains, there is most certainly a role for statement formalization, and we have provided for this as an important interface in our model.

We have also attempted to show how the model not only meets various core use cases, but that it can be applied across a spectrum of complexity, from very simple annotations of the kind researchers make nearly every day using reference management and annotation software, to complex curated knowledgebases. This is a basic requirement for success in adoption: you do not have to buy the whole package. There is a simple entry point beyond which you need go no further. But you may go on, if you so choose.

We assume that implementation will be in various kinds of software environments in which the actual detailed construction of micropublications is built into useful activities,

where the formal instantiations of micropublications are constructed internally and behind the scenes – but may be shared at will. Our research group has built a micropublications capability into the Domeo web annotation platform as a first step, and an evaluation of the model in this context will be presented in a forthcoming article.

As a general point, we expect value to accrue in the use of this model by enabling interoperability and incremental value addition through annotation, by users at various points in the scientific communication's ecosystem, who receive significant added utility in return for any additional work they put into modeling activities. In use cases such as bibliographic management tools, much of the required modeling work is already done in everyday use of the existing tools. What is added (claim identification in cited works and specification of representative or holotype claims), is certainly going to be useful to the users who implement it.

The most promising approaches overall, will generate micropublications as standardized annotation metadata that can be exchanged and accumulated between applications.

As pointed out in [88], when annotation metadata is published using with the W3C Open Annotation Model [87] and includes semantic tags on the statements, it becomes a first-class object on the Web, and can be published, if desired, as linked data. This will not always be desirable due to privacy concerns for certain kinds of annotations and comments, and/or licensing issues. But in many cases we believe it would be useful.

We hope that other researchers will see the utility of one or more elements of this model for their own use cases of choice, and apply it. Developers and architects who wish to explore further uses of this model, including modifications and new applications, are invited to contact and discuss with the authors.

Availability

The micropublications ontology may be accessed at http://purl.org/mp/. It is made available by the copyright holder, Massachusetts General Hospital, for use under terms of the W3C open source license (http://www.w3.org/Consortium/Legal/2002/copyright-software-20021231).

Endnotes

[a]Class names in the model are always capitalized, and are italicized on first occurrence. Predicate names are always in lower case and always italicized. When a term such as "claim", "statement", or "supports", is not being used as a formal class or predicate name, it is not capitalized or italicized.

[b]W.V.O. Quine criticized Strawson's formulation for similar reasons [131].

[c]However, note that the author does not actually say "these mice are a good model of AD".

[d]All three of the current authors participated in the work of the W3C group, and the second author co-chaired it; overall the W3C Open Annotation Community Group has had the participation of 55 institutions; with 97 participating individuals and representatives at the time of writing.

Additional files

Additional file 1: OWL Vocabulary and RDF Examples.

Additional file 2: Class, Predicate and Rule Definitions for Micropublications.

Additional file 3: Relationship of Micropublications to the SWAN Model.

Author details
[1]Department of Neurology, Massachusetts General Hospital, 55 Fruit Street, Boston, MA 02114, USA. [2]Harvard Medical School, 25 Shattuck Street, Boston, MA 02115, USA. [3]School of Computer Science, University of Manchester, Oxford Road, Manchester M13 9PL, UK.

References

1. Hunter L, Cohen KB: **Biomedical language processing: what's beyond PubMed?** *Mol Cell* 2006, **21**(5):589–594.
2. Krallinger M, Vazquez M, Leitner F, Valencia A: *Results of the BioCreative III (Interaction) Article Classification Task.* Bethesda, MD: BioCreative III Workshop; 2010:17–23.
3. Greenberg SA: **How citation distortions create unfounded authority: analysis of a citation network.** *Br Med J* 2009, **339**:b2680.

4. Greenberg SA: **Understanding belief using citation networks.** *J Eval Clin Pract* 2011, **17**(2):389–393.
5. Lawless J: **The bad science scandal: how fact-fabrication is damaging UK's global name for research.** In *The Independent.* London: Independent Print Ltd; 2013 [http://www.independent.co.uk/news/science/the-bad-science-scandal-how-factfabrication-is-damaging-uks-global-name-for-research-8660929.html].
6. Noorden RV: **Science publishing: The trouble with retractions.** *Nature* 2011, **478**:26–28.
7. Fang FC, Steen RG, Casadevall A: **Misconduct accounts for the majority of retracted scientific publications.** *Proc Natl Acad Sci* 2012, **109**(42):17028–17033.
8. Begley CG, Ellis LM: **Drug development: Raise standards for preclinical cancer research.** *Nature* 2012, **483**(7391):531–533.
9. Marcus A, Oransky I: **Bring On the Transparency Index.** In *The Scientist.* Midland, Ontario, CA: LabX Media Group; 2012.
10. Renear AH, Palmer CL: **Strategic reading, ontologies, and the future of scientific publishing.** *Science* 2009, **325**(5942):828–832.
11. Krallinger M, Morgan A, Smith L, Leitner F, Tanabe L, Wilbur J, Hirschman L, Valencia A: **Evaluation of text-mining systems for biology: overview of the Second BioCreative community challenge.** *Genome Biol* 2008, **9 Suppl 2**:S1.
12. Jelier R, Schuemie MJ, Veldhoven A, Dorssers LC, Jenster G, Kors JA: **Anni 2.0: a multipurpose text-mining tool for the life sciences.** *Genome Biol* 2008, **9**(6):R96.
13. Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Federhen S, Feolo M, Geer LY, Helmberg W, Kapustin Y, Landsman D, Lipman DJ, Lu Z, Madden TL, Madej T, Maglott DR, Marchler-Bauer A, Miller V, Mizrachi I, Ostell J, Panchenko A, Pruitt KD, Schuler GD, Sequeira E, Sherry ST, Shumway M, *et al*: **Database resources of the national center for biotechnology information.** *Nucleic Acids Res* 2010, **38**(suppl 1):D5–D16.
14. Clark T, Kinoshita J: **Alzforum and SWAN: the present and future of scientific web communities.** *Brief Bioinform* 2007, **8**(3):163–171.
15. Buckingham Shum S, Uren V, Gangmin L, Domingue J, Motta E: **Visualizing Internetworked Argumentation.** In *Visualizing Argumentation.* Edited by Kirschner PA, Buckingham Shum S. London: Springer-Verlag; 2003.
16. Shotton D: **Semantic Publishing: the coming revolution in scientific journal publishing.** *Learn Publishing* 2009, **22**(2):85–94.
17. Ciccarese P, Ocana M, Clark T: **Open semantic annotation of scientific publications with DOMEO.** *J Biomed Semantics* 2012, **3**(Suppl 1):S1.
18. Selventa Inc.: **Biological Expression Language V1.0 Overview.** [http://belframework.org/bel/overview]
19. Ciccarese P, Shotton D, Peroni S, Clark T: **CiTO+SWAN: The Web Semantics of Bibliographic References, Citations, Evidence and Discourse Relationships.** *J Biomed Semantics* 2014, **5**(4):295–311.
20. Ciccarese P, Wu E, Wong G, Ocana M, Kinoshita J, Ruttenberg A, Clark T: **The SWAN biomedical discourse ontology.** *J Biomed Inform* 2008, **41**(5):739–751.
21. **SWAN/SIOC: alignment between the SWAN and SIOC ontologies. W3C interest group note 20 October 2009.** [http://www.w3.org/TR/hcls-swansioc/].
22. Mons B, van Haagen H, Chichester C, Hoen PB, den Dunnen JT, van Ommen G, van Mulligen E, Singh B, Hooft R, Roos M, Hammond J, Kiesel B, Giardine B, Velterop J, Groth P, Schultes E: **The value of data.** *Nat Genet* 2011, **43**(4):281–283.
23. Mons B, Velterop J: **Nano-Publication in the e-science era.** In *Workshop on Semantic Web Applications in Scientific Discourse (SWASD 2009) 2009; Washington DC, USA*; 2009. CEUR: [http://ceur-ws.org/Vol-523/Mons.pdf].
24. Groth P, Gibson A, Velterop J: **The Anatomy of a Nano-publication.** *Inform Serv Use* 2010, **30**(1):51–56.
25. Schultes E, Chichester C, Burger K, Kotoulas S, Loizou A, Tkachenko V, Waagmeester A, Askjaer S, Pettifer S, Harland L, Haupt C, Batchelor C, Vazquez M, Fernández JM, Saito J, Gibson A, Wich L: **The Open PHACTS Nanopublication Guidelines V1.8.** EU Innovative Medicines Initiative - Open PHACTS Project; 2012. RDF/Nanopublication Working Group; 2012 [http://www.openphacts.org/documents/publications/OpenPHACTS_Nanopublication_Guidlines_v1.8.1.pdf].
26. Williams AJ, Harland L, Groth P, Pettifer S, Chichester C, Willighagen EL, Evelo CT, Blomberg N, Ecker G, Goble C, Mons B: **Open PHACTS: semantic interoperability for drug discovery.** *Drug Discov Today* 2012, **17**(21-22):1188–1198.
27. Spilman P, Podlutskaya N, Hart MJ, Debnath J, Gorostiza O, Bredesen D, Richardson A, Strong R, Galvan V: **Inhibition of mTOR by rapamycin abolishes cognitive deficits and reduces amyloid-β levels in a mouse model of Alzheimer's disease.** *PLoS One* 2010, **5**(4):e9979.
28. Bacon F: *Francis Bacon: The New Organon.* Cambridge UK: Cambridge University Press; 2000.
29. Shapin S: **Pump and circumstance: Robert Boyle's literary technology.** *Soc Stud Sci* 1984, **14**(4):481–520.
30. Cox FEG: **History of human parasitology.** *Clin Microbiol Rev* 2002, **15**(2):595–612.
31. Steup M: **Epistemology.** In *The Stanford Encyclopedia of Philosophy (Winter 2012 Edition).* Edited by Zalta EN. Palo Alto CA: Stanford University; 2012.
32. Simkin MV, Roychowdhury VP: **Stochastic modeling of citation slips.** *Scientometrics* 2005, **62**(3):367–384.
33. Simkin MV, Roychowdhury VP: **A mathematical theory of citing.** *J Am Soc Inf Sci Technol* 2007, **58**(11):1661–1673.
34. Alsheikh-Ali AA, Qureshi W, Al-Mallah MH, Ioannidis JP: **Public availability of published research data in high-impact journals.** *PLoS One* 2011, **6**(9):e24357.
35. HYLAND K: **Writing without conviction? hedging in science research articles.** *Appl Linguist* 1996, **17**(4):433–454.
36. Gerstein M, Seringhaus M, Fields S: **Structured digital abstract makes text mining easy.** *Nature* 2007, **447**(7141):142.
37. Leitner F, Chatraryamontri A, Ceol A, Krallinger M, Licata L, Hirschman L, Cesareni G, Valencia A: **Enriching Publications with Structured Digital Abstracts: the human-machine experiment.** Bedford, MA: MITRE Corporation; 2012.
38. Holmes FL: **Argument and Narrative in Scientific Writing.** In *The Literary Structure of Scientific Argument: Historical Studies.* Edited by Dear P. Philadelphia: University of Pennsylvania Press; 1991:224.
39. Douven I: **Abduction.** In *The Stanford Encyclopedia of Philosophy (Spring 2011 Edition).* Edited by Zalta EN. Paolo Alto CA: Stanford University; 2011.
40. Lipton P: *Inference to the Best Explanation.* 2nd edition. New York: Routledge; 2004.
41. Harman GH: **The inference to the best explanation.** *Philos Rev* 1965, **74**(1):88–95.
42. Campos D: **On the distinction between Peirce's abduction and Lipton's Inference to the best explanation.** *Synthese* 2011, **180**(3):419–442.
43. Walton D, Reed C: **Argumentation schemes and defeasible inferences.** In *Working Notes of the ECAI '2002 Workshop on Computational Models of Natural Argument, 15th European Conference on AI, 2002.* Lyons, FR: European Coordinating Committee for Artificial Intelligence; 2002:45–55.
44. Toulmin SE: *The Uses of Argument.* Cambridge UK: Cambridge University Press; 2003.
45. Verheij B: **Evaluating arguments based on Toulmin's scheme.** *Argumentation* 2005, **19**(3):347–371.
46. Verheij B: **The Toulmin Argument Model in Artificial Intelligence. Or: how semi-formal, defeasible argumentation schemes creep into logic.** In *Argumentation in Artificial Intellgence.* Edited by Rahwan I, Simari G. Dordrecht: Springer; 2009.
47. Dung PM: **On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games.** *Artif Intell* 1995, **77**(2):321–357.
48. Besnard P, Hunter A: *Elements of Argumentation.* Cambridge MA, USA: MIT Press; 2008.
49. Cayrol C, Lagasquie-Schiex M-C: **Bipolar Abstract Argumentation Systems.** In *Argumentation in Artificial Intelligence.* Edited by Rahwan I, Simari GR. Dordrecht: Springer; 2009.
50. Baroni P, Cerutti F, Giacomin M, Simari GR: **Computational Models of Argument.** In *Frontiers in Artificial Intelligence and Applications.* Edited by Breuker J, Guarino N, Kok JN, Liu J, LdM R, Mizoguchi R, Musen M, Pal SK, Zhong N. Amsterdam: IOS Press; 2010.
51. Boella G, Gabbay DM, VDT L, Villata S: **Support in Abstract Argumentation.** In *Computational Models of Argument.* Edited by Baroni P, Cerutti F, Giacomin M, Simari GR. Amsterdam: IOS Press; 2010.
52. Reed C, Norman TJ (Eds): *Argumentation Machines.* Dordrecht: Kluwer Academic Publishers; 2010.
53. Rahwan I (Ed): *Argumentation in Artificial Intelligence.* Dordrecht: Springer; 2009.
54. Bench-Capon TJM, Dunne PE: **Argumentation in artificial intelligence.** *Artif Intell* 2007, **171**(10–15):619–641.
55. Alberts B, Kirschner MW, Tilghman S, Varmus H: **Rescuing US biomedical research from its systemic flaws.** *Proc Natl Acad Sci* 2014, **111**(16):5773–5777.

56. Gereffi G: **The Evolution of Global Value Chains in the Internet Era.** In *Electronic Commerce for Development.* Edited by Goldstein AE, O'Connor D. Paris FR: Organisation for Economic Co-Operation and Development; 2002.

57. Holsapple CW, Singh M: **The knowledge chain model: activities for competitiveness.** *Expert Syste Appl* 2001, **20**(1):77–98.

58. Lee CC, Yang J: **Knowledge value chain.** *J Manage Dev* 2000, **19**(9/10):783–793.

59. Prinz F, Schlange T, Asadullah K: **Believe it or not: how much can we rely on published data on potential drug targets?** *Nat Rev Drug Discov* 2011, **10**(9):712.

60. Landis SC, Amara SG, Asadullah K, Austin CP, Blumenstein R, Bradley EW, Crystal RG, Darnell RB, Ferrante RJ, Fillit H, Finkelstein R, Fisher M, Gendelman HE, Golub RM, Goudreau JL, Gross RA, Gubitz AK, Hesterlee SE, Howells DW, Huguenard J, Kelner K, Koroshetz W, Krainc D, Lazic SE, Levine MS, Macleod MR, McCall JM, Moxley Iii RT, Narasimhan K, Noble LJ, *et al*: **A call for transparent reporting to optimize the predictive value of preclinical research.** *Nature* 2012, **490**(7419):187–191.

61. Cohen KB, Hunter L: **A critical review of PASBio's argument structures for biomedical verbs.** *BMC Bioinformatics* 2006, **7** Suppl 3:S5.

62. Ramos M, Melo J, Albuquerque U: **Citation behavior in popular scientific papers: what is behind obscure citations? The case of ethnobotany.** *Scientometrics* 2012, **92**(3):711–719.

63. Lu Z: **PubMed and beyond: a survey of web tools for searching biomedical literature.** *Database* 2011, **2011**:baq036.

64. Wan S, Paris C, Dale R: **Supporting browsing-specific information needs: Introducing the Citation-Sensitive In-Browser Summariser.** *Web Semantics: Science, Services and Agents on the World Wide Web* 2010, **8**(2–3):196–202.

65. Velterop J: **Nanopublications: the future of coping with information overload.** *LOGOS: J World Book Commun* 2010, **21**(3–4):119–122.

66. Clark T, Goble C, Ciccarese P: **Discoveries and Anti-Discoveries on the Web of Argument and Data**. In *AAAI-14 Workshop on Discovery Informatics, July 27–31, 2014.* Quebec CA: American Association for Artificial Intelligence; 2014.

67. Clark T: **Next Generation Scientific Publishing and the Web of Data.** *Semantic Web J.* in press.

68. Arighi C, Lu Z, Krallinger M, Cohen K, Wilbur W, Valencia A, Hirschman L, Wu C: **Overview of the BioCreative III Workshop.** *BMC Bioinformatics* 2011, **12**(Suppl 8):S1.

69. Arighi C, Roberts P, Agarwal S, Bhattacharya S, Cesareni G, Chatr-aryamontri A, Clematide S, Gaudet P, Giglio M, Harrow I: **BioCreative III Interactive Task: an Overview.** *BMC Bioinformatics* 2011, **12**(Suppl 8):S4.

70. Arighi CN, Carterette B, Cohen KB, Krallinger M, Wilbur WJ, Fey P, Dodson R, Cooper L, Van Slyke CE, Dahdul W, Mabee P, Li D, Harris B, Gillespie M, Jimenez S, Roberts P, Matthews L, Becker K, Drabkin H, Bello S, Licata L, Chatr-aryamontri A, Schaeffer ML, Park J, Haendel M, Van Auken K, Li Y, Chan J, Muller HM, Cui H, *et al*: **An overview of the BioCreative 2012 Workshop Track III: interactive text mining task.** *Database (Oxford)* 2013, **2013**:bas056.

71. Arighi CN, Roberts PM, Agarwal S, Bhattacharya S, Cesareni G, Chatr-Aryamontri A, Clematide S, Gaudet P, Giglio MG, Harrow I, Huala E, Krallinger M, Leser U, Li D, Liu F, Lu Z, Maltais LJ, Okazaki N, Perfetto L, Rinaldi F, Saetre R, Salgado D, Srinivasan P, Thomas PE, Toldo L, Hirschman L, Wu CH: **BioCreative III interactive task: an overview.** *BMC Bioinformatics* 2011, **12** Suppl 8:S4.

72. Chatr-aryamontri A, Winter A, Perfetto L, Briganti L, Licata L, Iannuccelli M, Castagnoli L, Cesareni G, Tyers M: **Benchmarking of the 2010 BioCreative Challenge III Text Mining Competition by the BioGRID and MINT Interaction Databases.** *BMC Bioinformatics* 2011, **12**(Suppl 8):S8.

73. Hirschman L, Yeh A, Blaschke C, Valencia A: **Overview of BioCreAtIvE: critical assessment of information extraction for biology.** *BMC Bioinformatics* 2005, **6** Suppl 1:S1.

74. Krallinger M, Vazquez M, Leitner F, Salgado D, Chatr-aryamontri A, Winter A, Perfetto L, Briganti L, Licata L, Iannuccelli M: **The Protein-Protein Interaction tasks of BioCreative III: classication/ranking of articles and linking bio-ontology concepts to full text.** *BMC Bioinformatics* 2011, **12**(Suppl 8):S3.

75. Lan M, Su J: **Empirical investigations into full-text protein interaction Article Categorization Task (ACT) in the BioCreative II.5 Challenge.** *IEEE/ACM Trans Comput Biol Bioinform* 2010, **7**(3):421–427.

76. Leitner F, Chatr-aryamontri A, Mardis S, Ceol A, Krallinger M, Licata L, Hirschman L, Cesareni G, Valencia A: **The FEBS Letters/BioCreative II.5 experiment: making biological information accessible.** *Nat Biotechnol* 2009, **28**:897–899.

77. Leitner F, Chatr-aryamontri A, Mardis SA, Ceol A, Krallinger M, Licata L, Hirschman L, Cesareni G, Valencia A: **The FEBS Letters/BioCreative II.5 experiment: making biological information accessible.** *Nat Biotechnol* 2010, **28**(9):897–899.

78. Leitner F, Mardis S, Krallinger M, Cesareni G, Hirschman L, Valencia A: **An Overview of BioCreative II.5.** *IEEE/ACM Trans Comput Biol Bioinform* 2009, **7**:385–399.

79. Lu Z, Kao H, Wei C, Huang M, Liu J, Kuo C, Hsu C, Tsai R, Dai H, Okazaki N: **The gene normalization task in BioCreative III.** *BMC Bioinformatics* 2011.

80. Wu CH, Arighi CN, Cohen KB, Hirschman L, Krallinger M, Lu Z, Mattingly C, Valencia A, Wiegers TC, John Wilbur W: **BioCreative-2012 virtual issue.** *Database (Oxford)* 2012, **2012**:bas049.

81. Ciccarese P, Ocana M, Clark T: **DOMEO: A web-based tool for semantic annotation of online documents.** In *Bio Ontologies 2011.* Vienna, Austria: 2011 [http://bio-ontologies.knowledgeblog.org/297]

82. Kim J, Ohta T, Tateisi Y, Tsujii J: **GENIA corpus–semantically annotated corpus for bio-textmining.** *Bioinformatics* 2003, **19**(Suppl 1):i180–i182.

83. Rebholz-Schuhmann D, Yepes A, Van Mulligen E, Kang N, Kors J, Milward D, Corbett P, Buyko E, Beisswanger E, Hahn U: **CALBC silver standard corpus.** *J Bioinform Comput Biol* 2010, **8**:163–179.

84. Comeau D, Doğan R, Ciccarese P, Cohen K, Krallinger M, Leitner F, Lu Z, Peng Y, Torii M, Valencia A, Verspoor K, Wiegers T, Wu C, Wilbur W: **BioC: an interchange data format and tools for biomedical natural language processing.** *Database* 2013, **2013**:bat064.

85. Ciccarese P, Ocana M, Das S, Clark T: **AO: An open annotation ontology for science on the Web.** In *Bio Ontologies 2010: July 9–13, 2010.* Boston MA, USA: 2010 [http://www.w3.org/wiki/images/c/c4/AO_paper_Bio-Ontologies_2010_preprint.pdf]

86. Ciccarese P, Ocana M, Garcia Castro LJ, Das S, Clark T: **An open annotation ontology for science on web 3.0.** *J Biomed Semantics* 2011, **2** Suppl 2:S4.

87. Sanderson R, Ciccarese P, Sompel HV, Bradshaw S, Brickley D, Castro LJG, Clark T, Cole T, Desenne P, Gerber A, Isaac A, Jett J, Habing T, Haslhofer B, Hellmann S, Hunter J, Leeds R, Magliozzi A, Morris B, Morris P, Ossenbruggen J, Soiland-Reyes S, Smith J, Whaley D: **W3C Open Annotation Data Model, Community Draft, 08 February 2013.** W3C; 2013 [http://www.openannotation.org/spec/core/].

88. Ciccarese P, Soiland-Reyes S, Clark T: **Web annotation as a first-class object.** *IEEE Internet Comput* 2013, Nov/Dec 2013:71–75.

89. Schwitter R, Fuchs NE: **Attempto Controlled English (ACE): A Seemingly Informal Bridgehead in Formal Territory.** In *JICSLP'96.* Bonn, Germany: GMD STUDIEN; 1996.

90. Fuchs NE, Höfler S, Kaljurand K, Rinaldi F, Schneider G: **Attempto Controlled English: A Knowledge Representation Language Readable by Humans and Machines.** In *Reasoning Web, First International Summer School 2005, Msida, Malta, July 25–29, 2005, Revised Lectures, Lecture Notes in Computer Science 3564.* Edited by Eisinger N, Maluszynski J. Berlin Heidelberg: Springer; 2005.

91. Fuchs NE, Schwitter R: **Attempto Controlled English (ACE).** In *CLAW 96, First International Workshop on Controlled Language Applications.* Belgium: University of Leuven; 1996.

92. Sanderson R, Ciccarese P, Sompel HVD, Clark T, Cole T, Hunter J, Fraistat N: **Open Annotation Core Data Model Community Draft, 09 May 2012.** W3C; 2012 [http://www.openannotation.org/spec/core/]

93. de Waard A, Buckingham Shum S, Carusi A, Park J, Samwald M, Sándor Á: **Hypotheses, Evidence and Relationships: The HypER Approach for Representing Scientific Knowledge Claims.** In *Proceedings 8th International Semantic Web Conference, Workshop on Semantic Web Applications in Scientific Discourse*, Volume 523. Edited by Clark T, Luciano JS, Marshall MS, Prdu'hommeaux E, Stephens S. Berlin, 26 Oct 2009, Washingon DC: Lecture Notes in Computer Science, Springer Verlag; 2009.

94. Hunter A: **Hybrid argumentation systems for structured news reports.** *Knowl Eng Rev* 2001, **16**(4):295.

95. Aristotle: *Rhetoric.* Mineola NY: Dover Publications; 2004.

96. Carroll JJ, Bizer C, Hayes P, Stickler P: **Named Graphs, Provenance and Trust.** In *WWW '05 14th international conference on World Wide Web 2005.* Chiba, Japan: ACM; 2005:613–622.

97. **Semantic Web Publishing Vocabulary (SWP) User Manual.** 2006 [http://wifo5-03.informatik.uni-mannheim.de/bizer/wiqa/swp/SWP-UserManual.pdf]

98. Bandrowski AE, Cachat J, Li Y, Müller HM, Sternberg PW, Ciccarese P, Clark T, Marenco L, Wang R, Astakhov V, Grethe JS, Martone ME: **A hybrid human and machine resource curation pipeline for the Neuroscience Information Framework.** *Database* 2012, (2012):bas005.

99. Gardner D, Akil H, Ascoli GA, Bowden DM, Bug W, Donohue DE, Goldberg DH, Grafstein B, Grethe JS, Gupta A, Halavi M, Kennedy DN, Marenco L, Martone ME, Miller PL, Muller HM, Robert A, Shepherd GM, Sternberg PW, Van Essen DC, Williams RW: **The neuroscience information framework: a data and knowledge environment for neuroscience.** *Neuroinformatics* 2008, **6**(3):149–160.

100. Altman M, Andreev L, Diggory M, King G, Sone A, Verba S, Kiskis DL: **A digital library for the dissemination and replication of quantitative social science research.** *Soc Sci Comput Rev* 2001, **19**(4):458–470.

101. Altman M, King G: **A proposed standard for the scholarly citation of quantitative data.** *DLib Magazine* 2006, **13**(3/4): march2007-altman [http://www.dlib.org/dlib/march07/altman/03altman.html].

102. Hardy J, Selkoe DJ: **The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics.** *Science* 2002, **297**(5580):353–356.

103. de Calignon A, Polydoro M, Suárez-Calvet M, William C, Adamowicz David H, Kopeikina Kathy J, Pitstick R, Sahara N, Ashe Karen H, Carlson George A, Spires-Jones Tara L, Hyman Bradley T: **Propagation of Tau pathology in a model of early Alzheimer's disease.** *Neuron* 2012, **73**(4):685–697.

104. Pimplikar SW: **Reassessing the amyloid cascade hypothesis of Alzheimer's disease.** *Int J Biochem Cell Biol* 2009, **41**(6):1261–1268.

105. Armstrong RA: **The pathogenesis of Alzheimer's disease: a reevaluation of the; Amyloid cascade hypothesis.** *Int J Alzheimers Dis* 2011, **2011.**

106. Herrup K: **Alzheimer's Disease - Modernizing Concept, Biological Diagnosis and Therapy.** Edited by Hampel H, Carrillo MC. Basel Switzerland: S. Karger; 2012:194.

107. Bryan KJ, Lee H-g, Perry G, Smith MA, Casadesus G: **Transgenic Mouse Models of Alzheimer's Disease: Behavioral Testing and Considerations.** In *Methods of Behavior Analysis in Neuroscience*. 2nd edition. Edited by JJ B. Boca Raton FL, USA: CRC Press; 2009.

108. Gao Y, Kinoshita J, Wu E, Miller E, Lee R, Seaborne A, Cayzer S, Clark T: **SWAN: A distributed knowledge infrastructure for Alzheimer Disease research.** *Web Semantics: Science, Services and Agents on the World Wide Web* 2006, **4**(3):222–228.

109. Harrison DE, Strong R, Sharp ZD, Nelson JF, Astle CM, Flurkey K, Nadon NL, Wilkinson JE, Frenkel K, Carter CS, Pahor M, Javors MA, Fernandez E, Miller RA: **Rapamycin fed late in life extends lifespan in genetically heterogeneous mice.** *Nature* 2009, **460**(7253):392–395.

110. Mucke L, Masliah E, Yu G-Q, Mallory M, Rockenstein EM, Tatsuno G, Hu K, Kholodenko D, Johnson-Wood K, McConlogue L: **High-level neuronal expression of Aβ1–42 in wild-type human amyloid protein precursor transgenic mice: synaptotoxicity without plaque formation.** *J Neurosci* 2000, **20**(11):4050–4058.

111. Hsia AY, Masliah E, McConlogue L, Yu G-Q, Tatsuno G, Hu K, Kholodenko D, Malenka RC, Nicoll RA, Mucke L: **Plaque-independent disruption of neural circuits in Alzheimer's disease mouse models.** *Proc Natl Acad Sci* 1999, **96**(6):3228–3233.

112. D'Hooge R, De Deyn PP: **Applications of the Morris water maze in the study of learning and memory.** *Brain Res Rev* 2001, **36**(1):60–90.

113. Morris RGM: **Spatial localization does not require the presence of local cues.** *Learn Motiv* 1981, **12**(2):239–260.

114. Huang S, Houghton PJ: **Mechanisms of resistance to rapamycins.** *Drug Resist Updat* 2001, **4**(6):378–391.

115. Neuhaus P, Klupp J, Langrehr JM: **mTOR inhibitors: an overview.** *Liver Transpl* 2001, **7**(6):473–484.

116. Sabatini DM, Erdjument-Bromage H, Lui M, Tempst P, Snyder SH: **RAFT1: a mammalian protein that binds to FKBP12 in a rapamycin-dependent fashion and is homologous to yeast TORs.** *Cell* 1994, **78**(1):35–43.

117. Strawson PF: **On Referring.** *Mind* 1950, **59**(235):320–344.

118. Attwood TK, Kell DB, McDermott P, Marsh J, Pettifer SR, Thorne D: **Calling International Rescue: knowledge lost in literature and data landslide!** *Biochem J* 2009, **424**(3):317–333.

119. Attwood TK, Kell DB, McDermott P, Marsh J, Pettifer SR, Thorne D: **Utopia documents: linking scholarly literature with research data.** *Bioinformatics* 2010, **26**(18):i568–i574.

120. Clark T, Ciccarese P, Attwood T, Waard A, Pettifer S: **A Round-Trip to the Annotation Store: Open, Transferable Semantic Annotation of Biomedical Publications.** In *Beyond the PDF: January 19–21, 2011*. San Diego: University of California at San Diego; 2011.

121. Needham M, Mastaglia FL, Garlepp MJ: **Genetics of inclusion-body myositis.** *Muscle Nerve* 2007, **35**(5):549–561.

122. Askanas V, Engel WK: **New advances in the understanding of sporadic inclusion-body myositis and hereditary inclusion-body myopathies.** *Curr Opin Rheumatol* 1995, **7**(6):486–496.

123. Askanas V, Engel WK: **Sporadic inclusion-body myositis and its similarities to Alzheimer disease brain. Recent approaches to diagnosis and pathogenesis, and relation to aging.** *Scand J Rheumatol* 1998, **27**(6):389–405.

124. Askanas V, Engel WK, Alvarez RB: **Light and electron microscopic localization of beta-amyloid protein in muscle biopsies of patients with inclusion-body myositis.** *Am J Pathol* 1992, **141**(1):31–36.

125. Askanas V, Engel WK, Alvarez RB, McFerrin J, Broccolini A: **Novel immunolocalization of alpha-synuclein in human muscle of inclusion-body myositis, regenerating and necrotic muscle fibers, and at neuromuscular junctions.** *J Neuropathol Exp Neurol* 2000, **59**(7):592–598.

126. Askanas V, Engel WK, Bilak M, Alvarez RB, Selkoe DJ: **Twisted tubulofilaments of inclusion body myositis muscle resemble paired helical filaments of Alzheimer brain and contain hyperphosphorylated tau.** *Am J Pathol* 1994, **144**(1):177–187.

127. Askanas V, McFerrin J, Alvarez RB, Baque S, Engel WK: **Beta APP gene transfer into cultured human muscle induces inclusion-body myositis aspects.** *Neuroreport* 1997, **8**(9–10):2155–2158.

128. Askanas V, McFerrin J, Baque S, Alvarez RB, Sarkozi E, Engel WK: **Transfer of beta-amyloid precursor protein gene using adenovirus vector causes mitochondrial abnormalities in cultured normal human muscle.** *Proc Natl Acad Sci U S A* 1996, **93**(3):1314–1319.

129. Needham M, Mastaglia FL: **Inclusion body myositis: current pathogenetic concepts and diagnostic and therapeutic approaches.** *Lancet Neurol* 2007, **6**(7):620–631.

130. Gupta A, Bug W, Marenco L, Qian X, Condit C, Rangarajan A, Muller HM, Miller PL, Sanders B, Grethe JS, Astakhov V, Shepherd G, Sternberg PW, Martone ME: **Federated access to heterogeneous information resources in the Neuroscience Information Framework (NIF).** *Neuroinformatics* 2008, **6**(3):205–217.

131. Quine WV: **Mr. Strawson on logical theory.** *Mind* 1953, **62**(248):433–451.