Yoked Flows for Direct Representation of Scientific Research

Robert B. Allen

New York, NY, USA

ABSTRACT

We propose developing highly structured and interlocking, or yoked, descriptions for all aspects of scientific research reports. These structured descriptions would be based on rich standardized vocabularies. We use two principal sets of flows to provide such structured descriptions: (a) Research Design and Procedures; and (b) Hypotheses and Outcomes. The structured descriptions may also include the research question, threats to validity, and implications. We propose that the best way to capture and describe the structure of scientific research is by considering multiple flows which are yoked. The claims from the research are propositions and they can be coordinated in a knowledgebase. As an example, we examine Pasteur's study of germ theory and support interaction with the structured description of the study with a prototype graphical user interface. We also consider template structures for different parts of the research reports. Ultimately, structured research reports could be interwoven into structured and evolving digital-library knowledgebases.

KEYWORDS

Highly Structured Digital Library, Microworld, Research Designs, Simulation Space, Specific Comparisons, Transitional Propositions, Validity

1 INTRODUCTION

We have been exploring direct representation for scientific research reports. Direct representation proposes that entire research reports can and should be highly structured. Moreover, we propose that collections of research reports can be interwoven into a rich semantic knowledgebase.

1.1 Semantic Models

Causal models, whether explicit or implicit, are central to science. Scientific research articles would benefit from using highly structured models which support state changes and causal relations. We use "flows" as a generic term for sequences of transitions such as workflows. flowcharts, plans, mechanisms, and other causal sequences. Potentially, flows could be circular or have feedback loops.

Recently, we have focused on the comprehensive ontology SUMO [24] as the vocabulary for such models. One important feature of SUMO that distinguishes it from most other ontologies is the inclusion of rules. We also propose the adoption of object-oriented modeling [7] in place of traditional approaches to presenting and processing knowledgebases. We implement transitions between object states and apply linguistic models of "case roles" to describe them [9].

In previous work, we have proposed a broad framework for flows that can be applied across domains [6, 10]. We have conducted several studies describing mechanisms and systems with structured, semantic vocabularies. Building on the modeling techniques in [9], we describe steps toward developing a rich model-oriented knowledgebase to support science. We describe

Digital Infrastructures for Scholarly Content Objects (DISCO), September 2021

EMAIL: rba@boballen.info
ORCID: [0000-0002-4059-2587]

CEUR Workshop Proceedings (CEUR-WS.org)

policies for making these simulations plausible and useful. While our current work focuses on qualitative models, the approach should also support quantitative models.

1.2 Scientific Research Reports

There is a long tradition of research on scholarly publications (e.g., [1, 30, 32]). Structure has increasingly been added to descriptions of scientific research. Taken to the logical conclusion, we propose that research reports should be totally structured. Structured research reports have many advantages. For instance, they can support interactive interfaces for visualizing and exploring the relationships among interlocking flows. Visualization of flows is related to timeline visualizations (e.g., [2]).

Several types of flows are already widely used in science. Workflows are used to specify experimental procedures (e.g., [14]). Mechanisms are often central for describing complex phenomena [6, 11]. However, before our work, Research Designs (e.g., [28]) as distinct from Research Procedures have not been explored as structured flows.

Beyond describing aspects of workflows and research phenomena directly, other parts of science research reports make claims and generalizations about phenomena. These can be characterized as a type of discourse [1, 2, 13, 15, 19, 25, 30]. We agree with [21] that research inferences cannot be based simply on formal logic. Rather, they follow a preponderance of evidence and consistency with other results.

1.3 Pasteur's Germ Theory

Germ theory was a paradigm shift in biology. It was sparked by the development of the microscope and the resulting ability to see microbes. Louis Pasteur was a major proponent of germ theory, which was the notion that tiny organisms, invisible without the aid of a microscope, produced spoilage, fermentation, and some diseases.

One early controversy was whether microbes developed only from other microbes or whether they developed spontaneously. That is, whether existing organisms are needed to propagate new organisms and those existing organisms are carried by air currents. We focus on a version of Pasteur's classic experiments that explored spontaneous generation [23, 26]. Pasteur's experiments are generally regarded as pivotal in confirming the importance of microbes and how they propagate.[1]

1.4 Roadmap

In [3], we used Pasteur's germ theory experiments to illustrate the potential for applying direct representation to scientific research reports. In this paper, we return to that example and describe how several techniques proposed in our recent work can be implemented to produce unified scientific research reports.

Our primary goal is the development of the underlying modeling framework for the organization and application of scientific knowledge. These models emphasize causal relationships (rather than classification) so we focus on what might be called transitional propositions. We also describe an interface for interacting with the models.[2]

In short, we propose that the best way to capture and describe the structure of scientific research is by considering multiple flows which are yoked. The claims from the research are propositions that can be coordinated in a knowledgebase.

---

[1] While Pasteur's report was not as detailed as current research reports, it is straightforward and useful as an exemplar.

[2] At this point, we are not focused on inference or text mining.

2 STRUCTURED RESEARCH REPORTS

2.1 Models and Knowledge Structures

While science uses systematic manipulations and/or observations, it also crucially depends on models about the phenomena under investigation. We employ two major flows to capture these two aspects. The first describes the Research Design and Procedure while the second describes the Hypothesis (i.e., what might happen) and Outcomes (i.e., what did happen).

Research microworlds are where the manipulations come together with the phenomenon under investigation. States and state changes are useful (at least implicitly) for models that describe dynamic environments. Some states are based on the properties of objects. Other states are based on the relationship among objects (e.g., an object is "trapped"). Sealing a flask is a complex action that achieves a state of separation. Breaking the flask is a way to unseal it and instantiate a new state in which the external air can move into the flask. Many research activities are workflows that involve multiple steps and interlock with other flows [9, 10].

While much of science is concerned with developing general principles, sciences such as geology and astronomy, as well as clinical medicine, deal more with particulars. Reasoned models can be developed for either general (abstract) principles or instances.

2.2 Creating a Research Space

Traditional research papers follow the IMRD (Introduction, Methods, Results, Discussion)[3] [32] framework. Swales [32] described the purpose of the Introduction of a research report as "creating a research space" (CARS). This includes defining a Research Question, Research Motivation, and Hypotheses.

Addressing the Research Question is the immediate goal of the research. Typically, it involves determining the existence, properties, mechanisms, processes, or applications associated with an entity or phenomenon. In some cases, the goal may simply be the replication of other research or addressing some criticisms that were raised about prior work. In this paper, we require that Research Questions can be answered with structured propositions.[4]

Examples of the Research Motivation might be practical (e.g., to find cures for a disease) or simply to acquire knowledge. Either way, it is an axiom, a given representing a valuation. Additional statements may link the Research Question to the Research Motivation.

The researcher then establishes plausible hypotheses by considering the factors potentially relevant to the Research Question by referring to established principles and previous research.

2.3 Research Design and Procedures

Based on the hypotheses, a Strategy is determined. The Strategy consists of the Research Design and Research Procedure. The Design is an overall framework for obtaining valid results. Independent and dependent variables are key parts of the Design. Typically, one of the hypotheses proposes some causal relationship between the independent and dependent variables. The independent variable may be manipulated either directly or indirectly. In natural experiments, the researcher identifies a natural event that creates conditions suitable for the research. These may include cases from natural science, social science [8], and medical science (e.g., the effects of smoking on cancer). In field and laboratory experiments, the researcher takes specific actions to manipulate the test environment.

Standard Research Designs are so entrenched in some fields that many researchers are unaware of

---

[3] Some publications do not use the exact IMRD structure but usually follow some permutation of it.

[4] See https://plato.stanford.edu/entries/questions/

them. In other fields, a variety of research paradigms is used and their merits are debated. [28] is a well-known analysis of the issues with different research designs. It discusses a wide range of designs and provides a notation for describing them. Moreover, it compares the possible threats to valid inference using different research designs. While [28] is primarily based on field research with randomization such as is common in social science, it can and should be applied more generally.

It is highly desirable to have at least two conditions for comparison [28]. This is especially true when one group is a control group and there is randomization of participants across conditions. However, these recommendations are not followed when a second group is difficult or impossible to implement, or when the researcher believes that he/she knows about and has controlled for possible extraneous factors.

The Research Procedure is a script or plan for the researcher's actions. It applies methods and materials. Those are usually specific to the domain under investigation and may threaten the internal validity of the research if applied incorrectly.

2.4 Hypotheses and Microworlds

There is considerable controversy about the role of hypotheses in scientific research. In cases such as Pasteur's experiment discussed below, the hypotheses are sharply drawn and are associated with a distinct, although not necessarily fully understood, mechanism. However, in other cases, a hypothesis may be nothing more than a hunch.[5]

Our models are typically situated in a microworld[6] which is a spatial region that provides the context for the interaction of objects involved in the phenomenon under investigation [12]. The manipulations directly or indirectly change the state of the microworld and/or its contents. In other work (e.g., [7, 8, 10]) we allow complex microworlds; potentially, they could be subdivided and have different levels of temporal and spatial granularity.

2.5 Outcomes, Internal Validity, and Comparisons

As the research is conducted, the raw data can be structured and stored according to the semantic model. The data can be manipulated and workflows for data transformations and statistical analyses can be included[7] along with the massaged data.

Using the data, we can make comparisons across the flows. These comparisons are the basis for claims. Claims are propositions. They have a truth value that expresses a judgment or opinion about some aspect of the research (e.g., the causal relationship between the independent and dependent variables).

The primary comparison is set up by the Research Design. In Pasteur's study which we analyze below, the comparison is relatively simple. In other cases (e.g., [33]), the comparisons may involve complex objects and processes, and statistical tests that require additional flows.

Research must satisfy many constraints; many things can go wrong and invalidate the results. [28] identifies two major types of validity for research, internal and external validity[8]. Internal validity refers to problems with the Research Procedure and Methods, and whether they implemented the intended research conditions. The researcher may check on the effect of a novel

---

[5] Perhaps it would be better to use the term "potential explanation" rather than hypothesis. For example, in [4] we examined [33], a modern biology paper dealing with the protein pathway related to Wallerian Degeneration. That paper cast a wide net and tested hypotheses which seemed unlikely to be relevant.

[6] This term is adopted from object-oriented programming. In our applications, it may be more appropriate to call it a simulation space.

[7] These could follow the scripts of any of several statistical analysis packages, although a common interchange framework would be preferred.

[8] They also mention statistical conclusion validity and construct validity.

or tricky manipulation. Such checks on the manipulation would also be described with flows.

[28] lists potential threats to validity for each research design. Structured research reports should include specific structures for handling each of these issues. For instance, the outcome summary could have a list of hypotheses and challenges to their validity.

2.6 External Validity, Generalizations, and Explanations

External validity refers to the ability to generalize beyond the experiment. Some generalizations may be straightforward, but others would be based on conditions. [28] describes criteria for generalizations. Generalization may require referring to broader issues within the research area or in other areas.

We would like to model those broader contexts, but, in many cases, they are not currently part of any structured model base. Eventually, such a model base could be developed; until then we can sketch a temporary framework (see Section 3.4).

Explanations may simply state a general rule. They may also try to describe how the rule applies to a given situation. If pressed, a mechanism to support the rule might be given. For instance, if we were explaining why hot air balloons rise, we would assert the rule that "hot air rises" and then might go into a discussion of the molecular dynamics of gasses (see Section 4.1).

3 PASTEUR'S SPONTANEOUS GENERATION EXPERIMENT

3.1 Overview

Farmers have considerable interest in understanding and controlling fermentation. The results of Pasteur's studies [23, 26, 31] are of practical importance for endeavors such as dairy, wine, beer, tofu, and soy sauce making, and for controlling infectious disease. In [3], we used Pasteur's research to explore the possibilities for highly structured research reports. In this paper, we take another step toward realizing that goal. We consider one of a series of related experiments by Pasteur. Specifically, we develop flows and an interface for presenting a structured description of one of Pasteur's germ theory experiments.

Pasteur put a nutrient broth in two sets of flasks. He boiled the broth and then sealed the neck of the flasks. He observed the flasks and eventually broke the neck open on one set of them. The flasks that remained sealed did not show microbe growth, while the flasks with the broken necks did.

We separate two main streams of activity in describing the experiments. The first is the Researcher Activity Model, which is what the researcher does based on the Design and Procedure. The second is the Outcomes Model, which is what happens, or could happen, in the environment under investigation. Although we distinguish them, the two streams are closely interlinked or yoked.

We focus on modeling the microworld and frame the experiment as a research design with two conditions. In the first condition, broth-filled flasks are sealed and then observed indefinitely. In the second condition, the flasks are sealed but eventually broken to demonstrate that spoilage occurs once external air reaches the broth. The critical test, between the sealed and broken-neck flasks, is determined by the Research Design and the manipulations.[9]

By modern standards, Pasteur's description of the research is somewhat informal. For instance, although Pasteur mentions that he made multiple flasks, we do not know how many. For illustrative purposes, we have inferred details as needed to complete these examples.

---

[9] No systematic randomization was done and there was no statistically significant sample, but the control groups suggest the comparisons that can be made and that must be explicitly represented.

3.2 Prototype Interactive Interface

Figure 1 shows the Researcher Activity Model (left) and Outcomes Model (center). Each has two columns, for each of the two conditions. Also shown (right) are Actual results and the key comparison that indicates that H1 (Hypothesis1) is supported (lower right).

At the top of the interface, there are several options to control the features of the visualization. These include:

- Toggle Method Details: Presents detailed descriptions of the procedures.
- Toggle Model Details: Shows additional details of the models. Potentially, there would be unique IDs for each of the model entities and transitions and the ontological parents associated with each could be displayed [10].
- Threats [to validity] and Alternative Explanations
- Inferences, Related Research, Applications, and Commentary

The interface was implemented with Python using the Tk graphics library. Development is ongoing; the current version is tailored to the specific example and does not include all the features needed for other research reports.

Figure 1: Screenshot of our interactive interface. The Conditions (left) follow the Research Design (blue) and Research Procedures (maroon). The Hypothesis Models and expected results are shown in green. The main comparisons for the hypotheses are shown (far right) in red and the conclusion in gold.

Condition_0	Condition_1	Hypothesis0	Hypothesis1	Actual
Initialize	Initialize	(spontaneous generation)	(germ theory)	
Prepare Broth, Flask, Burner	Prepare Broth, Flask, Burner			
Manipulation_0	Manipulation_0			
Boil Broth	Boil Broth	Boiling Kills Microbes in Broth	Boiling Kills Microbes in Broth	
Seal Flask	Seal Flask	Block External Air from Entering	Block External Air	
Wait	Wait	Spontaneous Generation		
Observation_0	Observation_0	Microbial Growth	(No change)	(No change)
Check turbidity	Check turbidity	Turbid		
	Manipulation_1			
	Break Neck of Flask		Air with Microbes Reaches Broth	
	Wait			
	Observe and Repeat		Microbial Growth	
	Check turbidity		Turbid	Turbid
				H1 supported

3.3 Hypotheses and the Microworld Model

Because of the complex interaction of entities in the Microworld, developing the full hypothesis models required additions to our evolving ontology and modeling framework. While some air had live microbes suspended in it, the air in the sealed flask had no live microbes. Thus, the state of the air is correlated with its location and the history of that location.

We focus here on Hypothesis1 because it is much more specific than Hypothesis0. Hypothesis1 is justified by several claims:

- Microbes can be carried by air currents (0)
- Sealing the flask neck blocks outside air (1)
- Breaking flask neck allows outside air to enter (2)
- High temperatures kill microbes (3)
- Microbes feed in a nutrient medium (4)
- Microbes will reproduce given food and other suitable conditions (5)
- Metabolism by many microbes results in spoilage (6)

Earlier research by Pasteur had confirmed (0). The other claims are largely consistent with common sense, though they could be tested more systematically as needed. However, even with extended testing, it is difficult to make an unassailable case [21].

A full executable flow model for Hypothesis1 would be analogous to the flow model in [10]. Note that a model for Hypothesis1 would need to include models of airflow in the microworld, ad hoc subregions for the air in the flasks, and multi-granular models that describe transitions of individual microbes as well as collections of microbes.

Because they are yoked, any execution of the Hypothses1 model should execute the parallel Researcher Activity model.

3.4 Outcomes

Raw data and inferences based on those data can be collected and organized according to the models described here. In Pasteur's study, the key observation is whether spoilage develops once the flask neck is broken and microbe-laden air can enter. That is, the critical test for the Pasteur study supports Hypothesis1, that the living microbes carried by air currents lead to spoilage.[10] We did not model the Actual Outcomes in this case, but we could have because they could be different than the predictions of either of the Hypotheses.[11]

Based on accepting Hypothesis1, we can state two overall claims:

- Microbes do not develop spontaneously (7)
- Microbes develop from other microbes (8)
- Microbes develop only from other microbes of the same type (8')

(8') is a stronger version of (8). Initially, we might be less willing to accept it, but there are additional factors we might consider. For instance, flows for the reproductive processes of the microbes would provide support.

As noted earlier, the research outcomes need to satisfy both internal and external validity. Internal validity concerns what happened because of the experimental procedure. For instance, we could dismiss (*9)[12] based on the experience of farmers.

- A longer time is needed for spoilage to develop than was used (*9)

(*10) was proposed by Antoine Béchamp, one of Pasteur's critics. The claim was that sealing the flask prevented air with some "vital essence" from reaching the broth. In a follow-up study, Pasteur was able to dismiss this criticism with his well-known swan-neck flask experiment [26].

- Sealed air loses its vital essence (*10)

3.5 Generalizations

If we combine (6) with (8') we obtain (11).

- Spoilage due to fermentation can be minimized by controlling the presence of microbes (11)

This suggests the need for cleanliness to control contamination in the preparation of fermented products. Further, if we combine (3) with (11) we get (12), which is the basis of pasteurization.

- Spoilage due to fermentation can be controlled by heating the nutrient medium (12)

Joseph Lister generalized (3, 8', 13) to bacterial infections to study and promote the need for sterile surgery. Moreover, adding (14) yields (15).

- Bacteria are a type of microbe (13)
- Antiseptics kill bacteria (14)
- Bacterial infection can be minimized by antiseptics (15)

Given the importance of each of these inferences for humans, presumably additional work would be done. For instance, specific microbes and the

---

[10] We might note the initial observation, that the sealed flasks show no spoilage. For a more formal confirmation, we could conduct an additional study with a control group.

[11] In [33] the results demonstrated a type of protein binding that was not predicted by the authors.

[12] Following a convention in linguistics, the * indicates that the proposition is incorrect.

details of conditions for growth could be studied for each medium.

4 FUTURE WORK

4.1 Interface, Model, and Claims

The interface in Figure 1 is adequate for a straightforward experiment such as Pasteur's. However, many modern research papers are much more complex. For instance, [33] includes a description of developing a strain of Drosophila needed for the research. It then conducts a series of overlapping studies that makes a case for its conclusions although no one study provides a definitive test. In such a set of studies, a great many flows can be identified and modeled. The interface will need to be improved to provide better support for that complexity.

The model and interface should be able to reorganize the research report flows to fit the IMRD framework (see Section 2.2). An IMRD Methods section would include the Research Design, Procedure, Methods, and Materials. Each of these components should fit sub-structures or templates and be integrated into the overall IMRD framework.

Claims must be based on clear definitions [12]. We have proposed SUMO as an ontology. SUMO bases its rules on established definitions, but even these need to be expanded and refined.

Although we have related claims to natural language propositions, our structured approach does not require natural language. Moreover, the case roles may be more exactly defined for each transitional and its interaction with various objects.

In Section 2.6, we suggested that an explanation for a claim could present a rule and an underlying mechanism. There is a broader sense of explanations that they should engage users in a way that promotes understanding. For instance, an extension of Figure 1 could support graphical guided tours as explanations. More elaborate explanations may be tutorial and can be based on pedagogical techniques.

4.2 Knowledge Structures

Claims from research reports and general axioms could be collected into a comprehensive knowledgebase. Although comprehensive, such a knowledgebase would be fragmented, changing, and need to represent multiple viewpoints. Even for areas where there is considerable agreement, there are internally consistent areas of knowledge (e.g., Newtonian mechanics) that may be usefully modeled separately from their connection to broader models (e.g., quantum mechanics).

Any knowledgebase of claims will need a range of structured hedges to indicate the type of claim (conceptual/logical, empirical, etc.), level of confidence in the claim, and possible criticisms of it. We would use a preponderance-of-evidence criterion for the acceptance of claims.

To the extent that we want to do inference on these propositions, we will need to support both open and closed worlds [27] and temporal reasoning in a dynamic environment [18, 22, 29].

4.3 Services for the Scientific Knowledgebase

The knowledgebase of research reports and claims can be viewed as a digital library. In addition to structured research reports, the library could also include structured surveys and reviews. Such a library could be overlaid with services like those found in a text-based digital library such as metadata harvesting and search indexing. Because the contents are structured, daemons may be able to generate text versions of the reports and to identify redundancy and inconsistencies.

We emphasize propositions that make claims about state changes such as (8). In a knowledgebase these claims should be accompanied by metadata. The metadata should include basic details such as date and creator; they should also link to related claims. If the metadata are said to provide support for claims, the details of that support should be included.

There could be links across structured research reports that are analogous to citations [5]. Our focus is at the level of semantics rather than the characteristics of the documentation. Thus, rather than link authors, we link functionally and semantically related flows (e.g., about methods) that are shared across research reports. In addition, measures analogous to citation metrics and alt-metrics could be developed for the strength of claims and the coherence of the knowledgebase [16].

Finally, (structured) annotations and commentary could be added. And administrative and editorial policies should be developed for managing the collection.

4.4 Envoi

We have proposed using yoked flows to manage the complexity of scientific research reports and have presented a prototype of a user interface for exploring those flows.

In addition, we have discussed issues for how claims from empirical scientific research can be collected and coordinated. The discovery and evaluation of causal claims are common to other scientific paradigms [20]. While those other paradigms may have different procedures than empirical research, they are also based on flows. Even if we distinguish classification (e.g., identifying different types of microbes) as a scientific research activity, flows are still used and could be modeled.

While we have pointed out some promising directions, there is still challenging work to be done in populating and organizing a large knowledgebase of credible propositions.

5 REFERENCES

1 R.B. Allen, Highly structured scientific publications. ACM/IEEE Joint Conference on Digital Libraries, 472, 2007, doi: 10.1145/1255175.1255271

2 R.B. Allen, Visualization, causation, and history, iConference, 2011, doi: 10.1145/1940761.1940835

3 R.B. Allen, Model-oriented scientific research reports, D-Lib Magazine, 2011 doi: 10.1045/may2011-allen

4 R.B. Allen, Rich linking in a digital library of full-text scientific research reports, Columbia University Research Data Symposium, 2013, RDStext.pdf

5 R.B. Allen, Rich semantic models and knowledgebases for highly-structured scientific communication, 2017, arXiv: 1708.08423

6 R.B. Allen, Issues for using semantic modeling to represent mechanisms, 2018, arXiv: 1812.11431

7 R.B. Allen, Definitions and semantic simulations based on object-oriented analysis and modeling, 2019, adXiv: 1912.13186

8 R.B. Allen, Metadata for social science datasets, In: Rich Search and Discovery for Research Datasets: Building the Next Generation of Scholarly Infrastructure, Edited by J.I. Lane, I. Mulvany, P. Nathan, Sage Publishing, 2020, PDF

9 R.B. Allen, Semantic modeling with SUMO, 2020, arXiv: 2012.15835

10 R.B. Allen, YM. Chu, Semantic models of pottery making, Pacific Neighborhood Consortium, 2021 in press

11 W. Bechtel, A. Abrahamsen, Explanation: A mechanist alternative. In Special Issue: Mechanisms in Biology. Studies in History and Philosophy of Science Part C Studies in History and Philosophy of Biological and Biomedical Sciences, 36(2), 421-41, 2005, doi: 10.1016/j.shpsc.2005.03.010

12 J.J. Cimino, Desiderata for controlled medical vocabularies in the twenty-first century, 1998, Methods of Information in Medicine, 1998, 37(4-5), 394–403, PMC: 3415631

13 T. Clark, P.N. Ciccarese, C.A. Goble, Micropublications: A semantic model for claims, evidence, arguments, and annotations in biomedical communications, Journal of Biomedical Semantics, 2014, doi: 10.1186/2041-1480-5-28

14 D. de Roure, C. Goble, R. Stevens, The design and realisation of the myexperiment virtual research environment for social sharing of workflows. Future Generation Computer Systems 25, 2009, 561-567. doi: 10.1016/j.future.2008.06.010

15 A. deWaard, 2010, From proteins to fairytales: Directions in Semantic Publishing, Intelligent Systems, IEEE, 25(2), 83-88, 2010, doi: 10.1109/MIS.2010.49

16 M. Friendman, Explanation and scientific understanding, Journal of Philosophy, 71(1), 1974, 5-19

17 R. Furuta, F.M. Shipman, CC. Marshall, D. Brenner, H-W. Hsieh, Hypertext paths and the world-wide web: experiences with Walden's paths, ACM Hypertext, 1997, doi: 10.1145/267437.267455

18 A. Galton, Spatial and temporal knowledge representation, Earth Science Informatics, 2, 169–187, 2019, doi: 10.1007/s12145-009-0027-6

19 N.L. Green, Toward mining scientific discourse using argumentation schemes, Argument & Computation 9, 121–135, 2019, doi: 10.3233/AAC-18003

20 T. Hey, S. Tansley, K. Tolle (eds.), The Fourth Paradigm: Data-Intensive Scientific Discovery, Microsoft Research, 2009

21 .I. Lakatos, A. Musgrave, Criticism and the growth of knowledge, Proceedings of the International Colloquium in the Philosophy of Science, London, V.4, Cambridge University Press, 1965

22 E.T. Mueller, Commonsense Reasoning: An Event Calculus Approach, Morgan Kaufmann/Elsevier, 2014, ISBN-13: 978-0128014165

23 L Pasteur, Sur les corpuscules organisés qui existent dans l'atmosphère: examen de la doctrine des générations spontanées: leçon professée à la Sociéte chimique de Paris le 19 mai 1861, par L. Pasteur.

24 A. Pease, Ontology: A Practical Guide, Articulate Software Press, Angwin, CA, 2011

25 M. Pera, The Discourses of Science, Translated: C. Botsford, University of Chicago Press, 1994

26 J.R. Porter, Louis Pasteur: Achievements and disappointments: 1861, Bacteriological Reviews, 25(4), 389-403, 1961, doi: 10.1128/br.25.4.389-403.1961

27 A. Rector, J.Rogers, A. Taweel, Models and inference methods for clinical systems: A principled approach, Studies in Health Technology and Informatics, 2004, 107(Pt 1), 79-83, PMID: 15360779

28 W.R. Shadish, T.D. Cook, D.T., Campbell, Experimental and Quasi-Experiment Designs for Generalized Causal Inference, Houghton, Mifflin, Boston, 2002,

29 M. Shanahan, The Event Calculus Explained, LNAI (1600) 409-430, 1999, Citeseer: 10.1.1.596.7046

30 S.B. Shum, E, Motta, J. Domingue. ScholOnto: An ontology-based digital library server for research documents and discourse. International Journal on Digital Libraries, 3, 237–248, 2000, doi: 10.1007/s007990000034

31 D.E. Stokes, Pasteur's Quadrant – Basic Science and Technological Innovation. Brookings, 1997, ISBN 9780815781776

32 J.M. Swales, 2004, Research Genres: Explorations and Applications. Cambridge University Press, Cambridge UK, 2004

33 R.G. Zhai, Y. Cao, P. R. Hiesinger, Y. Zhou, S.Q. Mehta, K.K. Schulze, P. Verstreken, H.J. Bellen, Drosophila NMNAT maintains neural integrity independent of its NAD synthesis activity. PLOS Biology, 4, 2006, doi: 10.1371/journal.pbio.0040416