Understanding Semantic Search on Scientific Repositories: Steps towards Meaningful Findability

Thiago Gottardi[1][0000−0003−0432−2034], Claudia Bauzer Medeiros[1][0000−0003−1908−4753], and Julio Cesar Dos Reis[1][0000−0002−9545−2098]

Institute of Computing, University of Campinas – UNICAMP, Campinas, Brazil
{gottardi,cmbm,jreis}@ic.unicamp.br

Abstract. Open Science has become one of the most important movements for leveraging scientific collaboration through sharing and reuse of research outputs, such as publications, data, software and methods, which are stored and preserved in public repositories. The emergence of FAIR principles as mandatory for effective openness has made it clear that it is not enough to publish such outputs – "findability" (our focus) is a key issue. It can be leveraged by good documentation and, e.g., PIDs (persistent identifiers), but search mechanisms have yet to adapt to this new context. Many such mechanisms – in particular semantic search – have risen as a means towards findability. Nevertheless, implementing these mechanisms and integrating them into scientific repositories presents many challenges. This paper presents a systematic literature review of research efforts on mechanisms designed to support search for research outputs – publications, data and processes. Our analysis is based on processing the entire collection of papers stored in Scopus, IEEE Xplore and arXiV. We identified 299 papers related to semantic search in scientific repositories. Their analysis provided a categorization of existing literature and unexplored gaps, pointing out new research challenges for semantic search mechanisms to support Open Science.

Keywords: Open Science · Semantic Search · Scientific collaboration.

1 Introduction

The sharing of research outputs has become a key enabler for Open Science [20], thereby enabling advancement of science through reuse of such outputs. Open Science relies on a combination of three major factors: open publications, open data, and open processes and methods, all made available in public repositories.

A major obstacle for the effective reuse is findability of data - and thus the institution of FAIR principles for data sharing and reuse [19], extensible to papers and processes (which include, among others, software and workflows). We identified that these three factors, together with authors, constitute the four most important parameters considered by search mechanisms. To avoid constant enumeration of these four parameters, we simply refer to them as *classes*.

Search mechanisms are cumbersome, and often require lengthy efforts to identify artifacts of interest. Several research solutions were proposed to alleviate the search process – such as the use of metadata standards, consensual vocabularies, or annotations. Semantic search mechanisms have risen as a means to improve the quality of search results. Many mechanisms vary widely in approaches and purposes. Our main concern is with semantic search mechanisms that serve Open Science purposes, namely *supporting search for scientific papers, data, and software in public repositories.* Besides these purposes, we also discuss the availability of search for authors and their affiliations.

Which search mechanisms, however, are best suited to ensure meaningful findability? Indeed, few studies are concerned with literature review on semantic search issues. To the best of our knowledge, there are no systematic literature reviews on the context of semantic search and its integration to scientific repositories; rather, surveys cover associated issues. For instance, Xu *et al.* [22] presented a study on semantic search by providing a survey on schemas for metadata associated to scientific publishing; Zhang *et al.* [23] studied approaches to identify the requirements for metadata search in the context of scientific data management. The work of Karimi *et al.* [8] analysed different approaches that employ thesauri and ontologies for semantic search. As examples of loosely related systematic reviews, Nguyen and Gobinda [12] performed a systematic review to create a knowledge map of digital libraries, while Figueroa *et al.* [3] presented a review on the progress of linked data technology, Gacitua *et al.* [4] provided a systematic review on semantic web technologies and discussed how to apply them on data warehouses or other industrial uses.

We conducted a systematic literature review to identify mechanisms that promote findability. A preliminary version of this work was published as a short paper [7]. This review focuses on semantic search mechanisms on public repositories containing papers, data or processes – from now on called *scientific repositories.* A systematic mapping is a method that allows to present empirical data from a broad subject of interest [14], thereby structuring a research area. After processing all documents from IEEE Xplore, Scopus and arXiV, we identified 478 relevant documents (of which 324 are unique). We provide a quantitative summarization, and a qualitative categorization and descriptions of the objectives and class of objects employed in the corresponding approaches. Our results indicate that most semantic search approaches lack several factors to fully meet findability – *e.g.*, flexibility in search parameters, or support to multiple domains. Our analysis points out that there are still many research challenges on the use, design and implementation of mechanisms for semantic search on open scientific repositories – in particular, on how they can be enhanced to provide a more meaningful range of results. As such, we provide insights into steps towards semantic search efforts, in particular to meet the demands of the open science movement. We thus present two major contributions – the systematic review itself, and its discussion; and the presentation of a few major open problems concerning semantic search mechanisms for open science.

2 Applying the Systematic Mapping Methodology

Our literature review follows the structure of a systematic mapping [14] and was executed according to guidelines [10]. These guidelines involve three sequential phases: (1) Planning; (2) Conducting and (3) Reporting. This section briefly outlines the Systematic Mapping Methodology, and how we applied it to analyze publications on semantic search mechanisms.

Planning, the first phase of the systematic mapping process, produces “Protocol”. Our Protocol appears in [6]; it specifies among others the Sources to be used in the review - namely, Scopus, IEEE Xplore and arXiV. Table 1 contains a subset of our protocol items.

Source selection criteria are refined by Inclusion (I1 to I3) and Exclusion (E1 to E3 criteria). I1 was planned to select all papers that involve any kind of search or query of databases. I2 was planned to include papers discussing any kind of integration scheme, *e.g.*, integration to databases or integration of different datasets. I3 involves selecting studies that present any kind of semantic mapping approach, *e.g.*, annotation, metadata or ontologies that are used to map (associate) documents to semantic predicates. We point out that we, on purpose, used in I1 and I2 terms that have more than one interpretation - “integration” and ”mapping”, so that we could select a larger set of papers to analyze.

Exclusion criteria are associated to studies that cannot be used for summarizing. In most cases, they were not scientific studies, *e.g.*, a call for papers. In other cases, they are unrelated studies that had the keywords employed as an unrelated meaning, *e.g.*, the search of something that is not scientific or does not use literature or database. We also excluded non primary studies, *i.e.*, studies that do not propose mechanisms, but are instead reviews themselves. Thus, they were not considered in our quantitative analysis.

The Conduction Phase is composed by the “Selection” and “Extraction” activities; The Selection phase defined which studies must be selected from the Sources, based on their titles and abstracts. The Extraction phase involved completely reading the documents for extracting data as planned by the Protocol.

Selection phase - Search for documents to review. The retrieval of documents of interest from the Sources was performed using different strings, which were used to identify three different categories of studies, using search strings. The first category concerned approaches that include any type of semantic search aspect, including semantics, ontologies, metadata or annotations. The second category comprised all types of data retrieval and search approaches, regardless of use of semantics or scientific repositories. The third category focused on synonyms for scientific, research and studies. “Research” was eventually removed for being far too common. “Study packing” was added since it has been used to refer to documentation of systematic reviews. Table 2 shows the search strings used in this phase, which were applied to title, abstract and keywords of all documents in each Source.

The initial search sessions on our three Sources were executed on February 17, 2020 and updated throughout August 11, 2020. While our focus is semantic

Table 1. Protocol Definition

Protocol Item	Item Description
Objective	Identify existing approaches to integrating semantic search mechanisms on scientific production.
Research Question	What are the approaches and techniques that perform semantic search on scientific production?
Intervention	Identify and categorize related primary studies.
Results	Quantitative data on frequency distribution within categories. Qualitative data on approaches that integrate semantic search on scientific databases.
Source Selection Criteria:	Source must be indexed studies on Computer Science, Mathematics or Engineering. Extraction of papers from source must allow Boolean operators. Source must be accessible by us.
Study Selection Criteria:	Inclusion I1 - Scientific Database Search approach; Inclusion I2 - Approach involves Integration; Inclusion I3 - Application of Semantic Mapping; Exclusion E1 – Not a valid document or inaccessible; Exclusion E2 – Unrelated to computing/databases; Exclusion E3 – Does not discuss search; Exclusion E4 – Not primary study.

search, our strategy for selecting studies for our systematic review includes other search approaches for completeness' sake on "findability" mechanisms.

After eliminating duplicated studies (e.g. that appeared in more than one Source), the review was conducted manually by exhaustively analyzing studies returned by each Source according to inclusion and exclusion criteria. Exclusion criteria rejected the study from the review, while studies meeting any inclusion criterion were included in the final set to be subsequently analyzed and summarized with respect to title, abstract and keywords.

Study Extraction. During the Extraction phase, all studies were qualitatively summarized by manually evaluating their full texts completely. An extraction form was filled manually for each study. The form contained the following fields:

Table 2. Search String Definition

Session	String
1	(("semantic search" OR "ontology search" OR "metadata search" OR "meta data search") AND ("scientific" OR "study pack" OR "study packing"))
2	(("semantic query" OR "ontology query" OR "metadata query" OR "meta data query") AND ("scientific" OR "study pack" OR "study packing"))
3	(("semantic information retrieval" OR "ontology information retrieval" OR "metadata information retrieval" OR "meta data information retrieval") AND ("scientific" OR "study pack" OR "study packing"))
4	(("semantic retrieval" OR "ontology retrieval" OR "metadata retrieval" OR "meta data retrieval") AND ("scientific" OR "study pack" OR "study packing"))

(I) Existence of Integrated Search (boolean); (II) Existence of Semantic Mapping (boolean); (III) Identified Software Architecture (nominal); (IV) Identified Objectives for Scientific Data (nominal); (V) Identified Class of Scientific Data (nominal);

Once the data from the Extraction phase was synthesized, we proceeded to the Reporting phase involving both quantitative and qualitative summarizing of the studies. The qualitative analysis was performed by writing textual descriptions for each study, clustering them according to their form data.

3 Results of the Systematic Mapping

First we show the sets of papers retrieved by the search strings, followed by an analysis of the results after each phase. Following the search sessions, we ended up with a total of 324 unique documents, of which 299 are valid studies.

Table 3. Results of the First Selection Phase

Input	I1 (Search)	E1 (No Document)	E2 (Unrelated)	E3 (No Search)	Output
299	280	9	1	4	276

During the selection phase, we selected studies that contain search approaches and excluded those that matched any exclusion criterion. This phase included studies with either integration or semantic mapping. For full description presented in this review, we cite the studies that include both criteria.

Table 4. Results of the Second Selection Phase

Input	I2 (Integration)	I3 (Semantic Mapping)	I2 ∩ I3	I2 ∪ I3	E4 (Non Primary)	Output
276	82	20	12	90	8	85

Table 5 presents the results of the Extraction phase that include both semantic mapping and integration. There were 12 studies in this category, however one study was excluded during extraction phase for being a non primary study and then moved to related work [13], causing the table to include 11 references. It cites the studies as the reference ("Ref") column, with the publication "Year", a short "Descriptive Summary", and their categorization, as follows: "**Integration**" refers the different ways in which each search approach integrated data: "Layer" stands for a semantic layer built on top of another database; "Multi" stands for the integration of multiple existing databases; "Existing" refers to annotating existing data to be enriched with semantics, *i.e.*, integration may be performed by an external semantic layer, or by integrating underlying databases, or indirectly via semantic annotations. "**Semantic Mapping**" includes the process executed to map semantics to data: either by "Manual" definitions or by "Auto" (automatic) definitions. We also identified if the approach is "Strict" or "Fuzzy" where applicable. "**Software Architecture**" cites the referenced software architectures. "**Object Class**" indicates the type of the data handled by the referred approach, according to the three main axes of Open Science: "S. Data" for scientific data; Papers, articles and other documents are referenced as"Document"; Methods, workflows, software and other processes for data handling are listed as "Process". "**Objective**" involves data usage intent of each approach. "Access" refers to data access, including search and retrieval; "Discover" refers to the discovery of new conclusions based on existing data; "Review" is the activity of surveying and aggregating data from other studies. Section 4 contains a further discussion of these categories.

Table 5. Results of the Extraction Phase

Ref.	Year	Descriptive Summary	Integration	Semantic Mapping	Software Architecture	Class	Objective
[21]	2007	A semantic model for annotating scientific data (material research) scattered over several databases.	Layer	Manual	Multiple Database	S. Data	Access
[16]	2008	Peer to Peer architecture for collaborative research.	Multi	Auto	Peer to Peer & Web	Document	Access & Review
[2]	2012	API for integrating biological data. Programmatic rules to map to existing ontology.	Multi & Layer	Auto & Strict	Programming Interface & Web	S. Data	Access & Discover
[9]	2013	A (semi)automatic crawler to build an ontology from document repositories that allows semantic search.	Layer & Existing	Auto	Multiple Database	S. Data	Access
[11]	2013	Dynamic semantic mapping between ontologies.	Layer & Multi	Auto	Multiple Database	S. Data	Access
[1]	2014	An ontology to map data inside databases.	Layer	Manual	Multiple Database	S. Data	Access
[24]	2014	A Middleware that adds semantic query capability for biomedical scientific data. Rules are written by the user.	Layer	Manual	Middleware	S. Data	Access
[5]	2017	A distributed architecture for software artifact catalog with semantic search capability. Mapping is crowdsourced.	Layer & Multi	Manual	Web	Process	Access
[17]	2017	An architecture proposal (not functional) for semantic annotation of existing scientific documents to allow semantic search.	Layer	Auto	Web	Document	Access
[15]	2019	Semantic mapping for Semantic web on top of relational databases.	Layer	Manual	Web & Multiple Database	Document	Access
[18]	2019	A semantic reasoning layer extension for Scala, which allows to integrate ontology search into programming.	Layer	Manual	Programming Interface	Data	Access

4 Discussion and Challenges towards Findability

4.1 Integrated Semantic Search

Semantic search has been applied to different scientific fields. The primary research question towards findability is related to the existence of "Integrated Semantic Search" in the literature. The term "integration" is used loosely, and can be found in many contexts. There were 77 total studies that involve some sort of "integration". We identified three different meanings for this term in the context of semantic search. The first meaning was how to connect multiple databases that include semantics with the intent to search them jointly – this was identified in 39 studies. The second was to take existing data and study how to add semantics to this data, identified in 34 studies. The third meaning relates to how to add a semantic layer to existing search engines, identified in another 34 studies. This semantic layer is closely related to semantic mapping (cf. Subsection 4.2). These integration concerns are related to the software architecture of the approaches (cf. Subsection 4.3).

Moreover, we were not able to find a generic proposal that was tested on multiple scientific fields – namely, an integrated approach to semantic search combining arbitrary domains. Rather, studies are motivated by or solely tested on a specific scientific domain, usually life sciences. Thus, a related question is: "how generic are the proposed mechanisms?". Many argue that, since their proposal is based on specific ontologies, changing the ontology would provide appropriate support to other domains. However, domain specificity hinders generality. The challenge is to balance between a domain-specific and a generic semantic search.

4.2 Semantic Mapping

Similar to "integration", there are different meanings of "Semantic Mapping". In general, it refers to metadata fields added to the actual data to enrich the data with semantic information. Our work identified 18 studies concerning semantic mapping, and identified three categories of this mapping. The first category, corresponding to 9 studies, is the "Manual Definition", in which metadata is manually specified by authors or curators, Since this represents a complex increase on work efforts, new approaches to automate these efforts were reported. In this sense, we identified a second category, which we named "Automatic Definition", in which metadata is added automatically by computers, presented in 9 studies. Automatic definitions also present challenges – *e.g.*, when algorithms add incorrect metadata. As part of efforts to address this issue, researchers created what we name "Fuzzy Mechanisms", which are variations of automatic metadata definitions, and which we identified in one study. Fuzzy mechanisms are those that use metadata to sort results by relevance, including loosely related - as opposed to "automatic definitions" in which only return directly related. We highlight that fuzzy definitions may lack precision. No identified study advocates "Strict Definitions", for example, the application of formal definitions to avoid ambiguity within the semantic search.

4.3 Software Architectures

Different software architectures have been adopted while designing integrated semantic search engines. We identified 77 studies that propose an integrated implementation. Most of the studies (24 in total) are based on multiple database composition, *i.e.*, the authors integrate several databases by implementing a single query system. This category of system appears in many situations, including large scale computing systems, *e.g.* clusters and grids, slowly been replaced by the emergence of cloud computing, which is represented by 5 papers.

A total of 20 studies indicate the use of web-based systems, often advocating that this implementation is adequate for the mainstream community. We also found many prototype proposals (reported by 12 papers); we could not check the actual architecture of these prototypes, since they were not described.

Semantic integration can be added as a layer to existing databases. Therefore, we expected studies suggesting middleware software solutions to support this kind of integration. However, only one study reported this attempt, which may indicate that this presents an implementation challenge to be followed up.

4.4 Objectives and Class Distributions

Class Distributions There are four main classes of search parameters declared in 64 studies: (a) Science Data: including text notes, spreadsheets, images, videos, recordings (41 studies); (b) Documents: including articles and theses (12 studies); (c) Processes: involving workflows; methods, hypotheses, comparison metrics, software (22 studies); and Author names and their affiliations (3 studies). Though the latter is not directly included in the three Open Science axes, it is a frequent parameter of search mechanisms. Considering the total number of processes and software repositories, we identified that a subset is not for scientific software (10 studies). Our results indicated that most papers only focus on a single object class. Indeed, out of 64 studies, only 13 deal with more than one data type and no study involved more than two classes. Thus, another research challenge is the design of (semantic) search mechanisms that allow combining distinct kinds of search parameters - documents, data, processes and authors.

Objective Distributions Regardless of the class of the employed data, the search mechanisms also had different objectives or goals, as reported on the study, involving 61 studies. The most common objective is to access the resulting data, moreover retrieving the results and to notify users when new results appear. The second most common objective is discovery of new conclusions that are not part of the original data submissions, including how to identify existing discovery aggregate data to identify and infer new conclusions (35 studies). A slightly less frequent objective is management, where existing data, documents and authors are registered and reported (23 studies). A less common objective, 3 studies focused on simulations; they may be used, for instance, to generate data for experiments and observations, validate data or extrapolate findings.

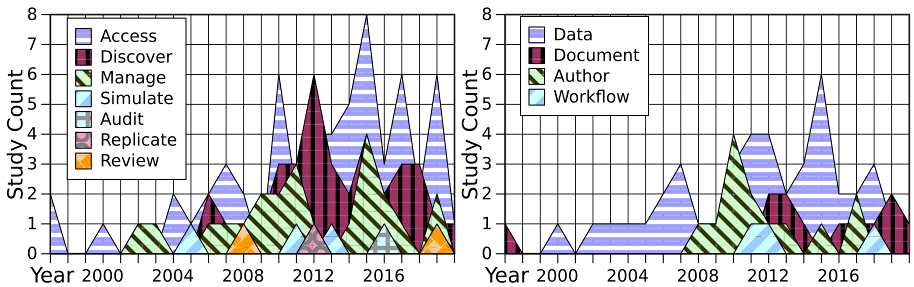

Fig. 1. Objective and Class Distributions over years

Another study focused on using the search mechanism for auditing data and conclusions; by using the collected data it is possible to identify the authors responsible for each claim, verify data, ensure correctness and detect frauds or corruption. The same study discussed reproduction or replication, where the experiments returned by the search should be reproduced or replicated to verify the findings. Finally, there were studies where the search was employed for supporting review efforts. Study reviews use existing documents and summarize them for creating new documents, aggregating their quantitative data and qualitative descriptions and comments into into a new (non-primary) study. Our review methods could eventually benefit from these search engines for collecting similar studies to be analyzed and summarized as new literature reviews.

Combinations and Opportunities of Classes and Objectives Table 6 shows the seven objectives and four object classes, resulting in combinations that employ the class (subject) with the action for an objective (verb). The table includes their frequency and descriptions; its columns are composed by the object class, while its rows are composed by objectives. Each cell in the middle contains the number of studies followed by its description. The descriptions are colored according to the number of identified studies: **green** identifies two or more studies, **orange** marks examples covered by a single study while **red** shows examples of no exisiting studies. Different combinations may indicate new opportunities for the usage of the given data, though some may not be feasible. Figure 1 includes a plot for the distribution of both class and usage objectives, showing how the number of studies vary from 1997 to 2020. It is worth reminding that the time this paper was written, the last year was not complete. These plots provide insights on periods in which objectives and object class appeared, *e.g.*, the increase on discovery studies or the rarity of workflow studies.

An analysis of the objectives shows future challenges, including cases that would benefit from semantic search. Some objectives identified in studies that are unrelated to semantic search – *e.g.*, studies concerned with prediction, which allow to “Predict” or estimate new data from existing data. Other studies advocated the support of “Data Export”, that allows users to take data results and explore them using software tools. Another challenge is to use semantic search

Table 6. Frequency and Descriptions for Objectives and Combination of Classes.

		Class			
		Scientific Data	Document	Process	Authors
Objective	Access	**29**: **Search, query, access, recommend and/or retrieve science data.**	**10**: **Search, query, access, recommend and/or retrieve papers, articles, journals, reports, magazines, etc.**	**8**: **Search, access, recommend and/or retrieve science data.**	**2**: **Search and find or recommend authors and related authors.**
	Discover	**22**: **Discover conclusions using aggregated science data.**	**4**: **Discover conclusions and related documents using existing documents.**	**7**: **Discover combined workflows.**	**1**: **Discover what authors collaborate on research efforts.**
	Manage	**13**: **Manage known science data, also their sources and bases.**	**2**: **Manage known document references/citations. Manage documents being written.**	**5**: **Manage known workflows and assess their usage.**	**1**: **Manage known authors, relationships, contributions and their roles.**
	Simulate	**3**: **Simulate experiments and compare against existing data for validation.**	**0**: **Simulate document publications and acceptance.**	**1**: **Simulate workflow usage and outcomes.**	**0**: **Simulate author contributions and outcomes.**
	Audit	**1**: **Audit data for validation and verification; protect from corruption and false data; blame manipulators.**	**0**: **Audit documents to verify authorship and protect documents from corruption.**	**0**: **Audit execution of workflows. Audit who can edit the workflow.**	**0**: **Audit roles and authorship to protect authors' curricula from corruption and false data.**
	Replicate	**1**: **Replicate studies based on existing science data and compare the outcomes.**	**0**: **Replicate (or plagiate) existing documents and their structures.**	**0**: **Replicate existing work-flows and compare their outcomes.**	**0**: **Plagiate author roles.**
	Review	**0**: **Review and compare data sets of science data to aggregate results.**	**1**: **Support for literature reviews.**	**0**: **Review work-flows and methods and compare their efficiency.**	**0**: **Review existing author roles and contributions.**

to find "Teaching" material for students. "Visualization" combined to semantic search could lead to better comprehension for both the semantic queries as well as the results from the semantic searches. There are also other objectives we identified after analyzing recent opportunities. For instance, there are no studies beyond the current data management tools that include strategic decisions for the future research efforts. Another opportunity is support the design of public "Policies" based on evidence. A completely missing objective we identified is the lack of semantic search to extract specific data and metadata from "Internal" content available in documents and data, *e.g.*, article sections or images.

5 Conclusion

Open Science relies on sharing research results, including publications, data and processes. Effective sharing requires findability, and for this we must understand research efforts on search mechanisms. Our work presented a systematic literature review on a promising approach – semantic search issues – analyzing and synthesizing 299 papers extracted from the entire collection of documents in IEEE Xplore, Scopus and arXiv.

This investigation presented both quantitative and qualitative results, providing new insights and pointing out open research issues to be addressed. Still, it is necessary to mention that there are threats to validity regarding the limited set of studies as well the construction validity of the search string, as well as the operation methods. These threats have been mitigated by exhaustively manual analysis of the results collected by search engines that aggregate data from more than one publisher.

We intend to perform new searches for completeness. We plan to proceed updating the review and continue analysing further studies to provide additional descriptive analyses. Ongoing work concerns extending the set of studies to be analyzed, by increasing the number of criteria and search engines.

The full set of results, including detailed methodology, graphic plots and analysis datasets appear in [6], which are also planned to receive incremental updates.

Acknowledgements This research is supported by FAPESP (#2019/19389-1, #2017/02325-5 and #2013/08293-7), as well as CNPq (#428459/2018-8 and #305110/2016-0). Opinions expressed here are solely the authors'.

References

1. Abburu, S., Golla, S.: A generic mapping method and tool to execute semantic queries on relational database. IJAER **9**(24), 29227–29238 (2014)
2. Deus, H., Zhao, J., McCusker, J., Fox, R., Prud'hommeaux, E., Malone, J., Das, S., Miller, M., Adamusiak, T., Rocca Serra, P., Marshall, M.: Translating standards into practice - one semantic web api for gene expression. In: SWAT4LS'11 (2012). https://doi.org/10.1145/2166896.2166900
3. Figueroa, C., Vagliano, I., Rocha, O.R., Morisio, M.: A systematic literature review of linked data-based recommender systems. Concurrency and Computation: Practice and Experience **27**(17), 4659–4684 (2015). https://doi.org/10.1002/cpe.3449, cpe.3449
4. Gacitua, R., Mazon, J.N., Cravero, A.: Using semantic web technologies in the development of data warehouses: A systematic mapping. WIREs Data Mining and Knowledge Discovery **9**(3), e1293 (2019). https://doi.org/10.1002/widm.1293
5. Gil, Y., Garijo, D., Mishra, S., Ratnakar, V.: Ontosoft: A distributed semantic registry for scientific software. In: IEEE e-Science. pp. 331–336 (2017). https://doi.org/10.1109/eScience.2016.7870916
6. Gottardi, T.: Documentation of meta-analysis for semantic search. Online; Updated October 2020 (2020), `http://tiny.cc/gottardi-semantic-review`

7. Gottardi, T., Medeiros, C.B., Reis, J.D.: Semantic search on scientific repositories: A systematic literature review. In: S. Brasileiro de Banco de Dados (SBBD) (2020)
8. Karimi, E., Babaei, M., Beheshti, M.: The study of semantic and ontological features of thesaurus and ontology-based information retrieval systems. Iranian Journal of Information Processing and Management **34**(4), 1579–1606 (2019)
9. Khattak, A.M., Ahmad, N., Mustafa, J., Pervez, Z., Latif, K., Lee, S.Y.: Context-aware search in dynamic repositories of digital documents. In: IEEE CSE (2013). https://doi.org/10.1109/CSE.2013.59
10. Kitchenham, B., Charters, S.: Guidelines for performing systematic literature reviews in software engineering. Tech. Rep. EBSE 2007-001, Keele University and Durham University Joint Report, UK (2007)
11. Luo, Y., Yu, Z., Zhuang, Y., Zheng, Z.: Dynamic mapping processing between global ontology and local ontologies in grid environment. Inf. Technol. J. (2013). https://doi.org/10.3923/itj.2013.2454.2459
12. Nguyen, S.H., Chowdhury, G.: Interpreting the knowledge map of digital library research (1990–2010). Journal of the American Society for Information Science and Technology **64**(6), 1235–1258 (2013). https://doi.org/10.1002/asi.22830
13. Niknia, M., Mirtaheri, S.: Mapping a decade of linked data progress through co-word analysis. Webology **12**(2) (2015)
14. Oakley, A., Gough, D., Oliver, S., Thomas, J.: The politics of evidence and methodology: lessons from the eppi-centre. Evidence & Policy **1**(1), 5–32 (2005)
15. Omar, Y.M.K., el Moneim, A.A., Mohamed, K.: Building a framework for mapping rdbms to rdf with semantic query capabilities. In: ITMS (2019). https://doi.org/10.1109/ITMS47855.2019.8940733
16. Pirrò, G., Ruffolo, M., Talia, D.: Advanced semantic search and retrieval in a collaborative peer-to-peer system. In: UPGRADE'08. pp. 65–71 (2008). https://doi.org/10.1145/1384209.1384222
17. Portilla Herrera, N.A., Gomez, F.L., Bucheli, V.A., Pabón, O.S.: Semantic annotation and retrieval of scientific documents in a big data environment. In: LACNEM 2017. pp. 1–6 (2017). https://doi.org/10.1049/ic.2017.0032
18. Seifer, P., Leinberger, M., Lämmel, R., Staab, S.: Semantic query integration with reason (2019). https://doi.org/10.22152/programming-journal.org/2019/3/13
19. Wilkinson, M., Dumontier, M., Aalbersberg, J., Appleton, G., et al.: The FAIR Guiding Principles for scientific data management and stewardship. Nature Data **3**(1) (2016). https://doi.org/10.1038/sdata.2016.18
20. Woelfle, M., Olliaro, P., Todd, M.H.: Open science is a research accelerator. Nature Chemistry **3**, 745–748 (October 2011). https://doi.org/10.1038/nchem.1149
21. Xiaoming, Z., Changjun, H., Qian, Z., Chongchong, Z.: Material scientific data integration for semantic grid. In: SKG 2007. pp. 414–417 (Oct 2007). https://doi.org/10.1109/SKG.2007.95
22. Xu, H., Sun, L., Zou, M., Meng, A.: A survey of scientific metadata schema. Applied Mechanics and Materials **411-414**, 349–352 (2013). https://doi.org/10.4028/www.scientific.net/AMM.411-414.349
23. Zhang, W., Byna, S., Niu, C., Chen, Y.: Exploring metadata search essentials for scientific data management. In: 2019 IEEE HiPC. pp. 83–92 (2019). https://doi.org/10.1109/HiPC.2019.00021
24. Zheng, S., Wang, F., Lu, J.: Enabling ontology based semantic queries in biomedical database systems. International Journal of Semantic Computing **8**(1), 67–83 (2014). https://doi.org/10.1142/S1793351X14500032