

Satellite Meeting: Grey Literature: Scholarly Communication in a Digital World

Date: Fri, August 23, 2019

*Location: National Library of Greece @ SNFCC (Stavros Niarchos Foundation Cultural Center),
Athens, Greece*

Metadata for Research Projects as Grey Node Hubs for Literature Repository

Jiuming Ji

Institute of Information Science and Technology, Library of East China University of Science & Technology, Shanghai, People Republic of China.

E-mail address: jjm@ecust.edu.cn

Weihong Luo

Jiangsu Zhongyang Data Technology Co., Ltd., Jiangsu Nantong, People Republic of China.

E-mail address: business@hiresearch.cn



Copyright © 2019 by Jiuming Ji and Weihong Luo. This work is made available under the terms of the Creative Commons Attribution 4.0 International License:

<http://creativecommons.org/licenses/by/4.0>

Abstract:

Compared with regular publications, Information of research projects is a type of grey literature. Fortunately, many commercial or non-commercial research project databases have been used as important information resources by many research-based universities. The problem is: on one hand, though some of these research project databases include the outcome and funding information of projects, many are settled separated from the other databases even from being displayed on the website of libraries; on the other hand, much grey literature has strong relationship with one or more research projects, such as meeting or conference information, research data sets, proposals and funding information, etc.. This paper introduces research projects as grey node hubs for literature repository, and metadata of research projects under this vision.

Keywords: grey literature, research project, metadata, grey hubs, literature repository.

Introduction

"A research project is a scientific endeavor to answer a research question." ¹This definition is displayed on the web of many universities for guiding students' research practice activities. There are two types of research projects based on being supported with funds or not. In this paper, the former type of research projects are referred as fund-granted and the latter is called self-granted research projects which usually being motivated by interesting of researchers themselves or being inspired by topics of conferences or journals. Most fund-granted research projects can only get supported after their proposals being assessed as eligibility and the funding information should be displayed in all outcome of it, so all research publications related to one certain research project can be collected easily.

It is exactly true that research lives of researchers, whether are supported by funds or not, are involved in life cycles of one or more projects, as shown in figure 1, research life of researchers. Thus, information about the research project is very important for researchers. On one hand, over 80% research articles, which are published in 2019 and indexed in Web of Science, were supported by funds. On the other hand, once fund proposal information was published by organizations, researchers would submit their application proposals. Some of these proposals were supported, others may be rejected. The researchers should work out the projects with the help of granting funds, some may insist to follow the ideas of proposals being rejected without the help of any funds. Some other researchers didn't care about funds, once they were interested in some research problems, which may not be supported by any funds. Thanks to contributions of publishers, libraries as well as funds sponsors, most research articles, books, reports, patents or technology secrets, etc have already been published or stored in repositories. Once been published, research publications or data may be cited by others, which makes linkages between one research project and the others. The link information among research projects can be mined for reviewing, assessment. More values of research project information will be described in related works.

Although some commercial or non-commercial research project databases have been used as important information resources by some research-based universities, some of them display the information of fund calling, others display the information of research outcome, for examples Grants.gov² and Proposal Central³ only display the information of fund calling, NSFC of China⁴ only displays the information of research outcome. Only a few of them display the information of granted, calling and research outcomes, such as Global scientific research project database of HiResearch (hereinafter referred to as HiResearch)⁵ and Scival

¹http://njms.rutgers.edu/departments/medicine/internal_medicine/documents/research.pdf

² <http://Grants.gov>

³ <https://proposalcentral.com/>

⁴ <http://or.nsf.gov.cn/>

⁵ <http://www.hiresearch.cn>

funding⁶. None of them displays the information about granted proposals. Those universities who don't have licenses of any research project databases, some of them only share notices of fund callings on web pages of library or university, others even notice none news of fund callings on web pages of library and university. Furthermore, most research project databases have not merged with databases of publications of research projects, most research paper or data databases only show less information about the related granted projects and no any out links to the research projects information pages. The key is that literature generated during the life cycles of most research projects can't be collected easily. It is not strange for the absence of research projects application or granted information in most articles which review the development of some certain research fields because of the absence of integrated research projects database, and the value of research projects information need to be fully utilized. This article aims to propose the links information of research project calling with proposals, granted and outcome. Compared with regular publications, information of research project is a type of grey literature, fortunately, any type of literature, both grey published and regular published, can have its metadata. The metadata is the best one for both nodes of literature and link between research projects to enhance the use of research project information. The object of this article is to design metadata of research projects as grey nodes for literature repositories.

Metadata Term

Although Priscilla Caplan (2003) referred catalog card schema of library, which has a long history, as a metadata schema, the term metadata comes from the field of computer science during later 1970s and early 1980s (John L. McCarthy, 1982), when statistics databases need more functions such as self-description (Becker R. A., 1978) and integrated metadata management (Teitel R. F., 1977).

John L. McCarthy from the University of California at Berkeley described metadata as data about data or systematic descriptive information about data content and organization that can be retrieved and manipulated, such as deleted, updated. This definition is very simple and abstract without constraints of scale and structure, which means it may be simple and unstructured, from man-readable typewritten narrative describing a data tape to machine-readable structured DBMS dictionary used to control multiple databases. Definitions and schemas of most types of objects, such as entities, attributes, data dictionaries, databases and its management, category sets, can all be described in some metadata, which contains names and aliases, labels and descriptive information, data derivations and qualities, security specifications, logical structure descriptions, access paths and linkage specifications, processing procedures, usage information, physical characteristics (John L. McCarthy, 1982). From then on, computer science researchers have concerned diversity problems from metadata knowledge descriptions (PMD Gray, 1988) to metadata management systems (L.

⁶ <http://www.funding.scival.com/>

Mark,1986) and to metadata description methods (Hong Yao, 2016) or repository (YS Joung, 2001), from using metadata to solve semantic conflicts (M. Siegel, 1991) to interaction between databases and systems (A. Alzobaidie, 1988). More and more metadata schemas of specific objects have been developed, such as hydrodynamic model data (Akms Islam,2006), thermal videos (Andras Hajdu, 2007), heritage collections (Athanasios D. Styliadis,2009), government data (Petr Kremen, 2019), manufacturing resource ontology models (Xiaobin Li, 2019), and digital image collections (Grace Therrell, 2019), etc.. These practices have already set the foundation for metadata schema of research projects for displaying, exchanging or harvesting.

Methodology

It's common that the most important problem of metadata is granularity, and it is authority controlling is also a vital problem. We used the method of investigation to resolve them. As shown in figure 1, the procedure contained three steps. First, we investigated the research projects information used by information resources. Second, we investigated the expectations of the researchers. Third, we analyzed the attributions collected from these two sources according to the usage expected by researchers.

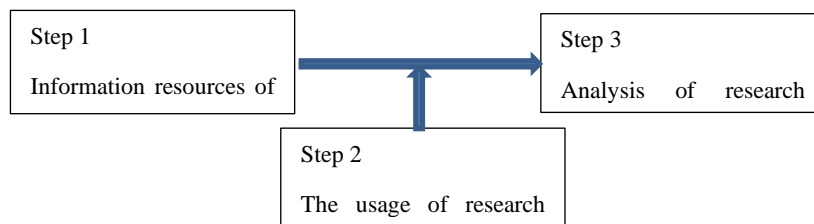


Figure 1 A procedure for creating the metadata of research projects

Investigation results

There are many research projects databases used by universities and other academical organizations all over the world, we chose some of them according to the investigation research finished by Qingfang He (2017). These databases mainly concern three statuses of research projects, such as forecast, granted and finished, more details in table 1. Pure is a research output repository portal product for research organizations which all have a very similar page of project information and stored the active or finished projects and some repository portals on it provide functions for projects searching. The other sites, such as Project Gate, Grants.gov, Proposal Center, PIVOT, etc, are also concentrated on research

projects information. All these sites provide the information of opportunities, and Scival Funding, Research Professional, CNKI and HiResearch provide all types of information of all state of research projects.

Table 1 metadata requirement investigation of research project

		forecast	Granted	finished
1	Pure/ ⁷ Elsever		√	
2	Project Gate	√	√	
3	Grants.gov ⁸	√		
4	Proposal central ⁹	√		
5	PIVOT/ProQest	√	√	
6	Scival Funding	√	√	
7	Foundation directory online	√		
8	Research Professional ¹⁰	√	√	√
9	SPIN	√		
10	CNKI	√	√	√
11	HiResearch ¹¹	√	√	√
12	Web of science			√

We investigated the attributes of research projects from bibliographies of these databases, web of science, acknowledgment of research articles and an inquiry survey, which also collected the usage needs of research project information. There are nearly 59, 38, 33 different attribute names or controlled value vocabularies for project opportunity information, granted projects, and finished projects from database or repository portal listed in table 1. For acknowledgment in articles or posters, there are only three attributes for granted projects, such as granted project codes or identities, funding organizations and its acronyms. The Grant.Gov contributes most vocabularies for forecast project information and the Research

⁷ <https://pure.qub.ac.uk/portal/en/projects/search.html>

⁸ <http://www.Grants.gov>

⁹ <https://proposalcentral.com/GrantOpportunities.asp>

¹⁰ [https://www.research.ed.ac.uk/portal/en/projects/a-live-pulse-yik-yak-for-understanding-teaching-learning-and-assessment-at-edinburgh\(fcc57661-8d8d-477e-9dc1-270fa812c3de\).html](https://www.research.ed.ac.uk/portal/en/projects/a-live-pulse-yik-yak-for-understanding-teaching-learning-and-assessment-at-edinburgh(fcc57661-8d8d-477e-9dc1-270fa812c3de).html)

¹¹ <http://www.hiresearch.cn>

Professional contributes most vocabularies for finished project information.

The survey was sent to Chinese researchers or students randomly and got 160 replies from 18.07% graduate students, 25.3% masters, 11.45% P.H.D candidates and 12.65% doctors, details are shown in figure 2 below. There were 129 interviewees having research experiences. As shown in figure 3, the percentage of senior and intermediate titles is 37.98% and 46.51% respectively. Furthermore, 72% of 129 researchers have heard of research projects. So, we selected the opinions of these 93 researchers to analyze the needs.

For the question "What information or attributes do you think should be included in metadata of research projects?" The top four concerned opportunity attributes were opportunity title, grantor, opportunity description, subject area, details shown in figure 4 below. As shown in figure 5, granted project title, controller name, the organization of controller, members name got popularity ratios over 73.12%. Although we only listed the literature type of information resources that related to some research projects, from the popularity figures, we can get some important hints about the link needs of the project and its dataset or patents, details shown in figure 6. From our investigation of research projects databases and repositories, we found that neither information of patents nor dataset being created or gathered to reveal their relations to the relevance projects.

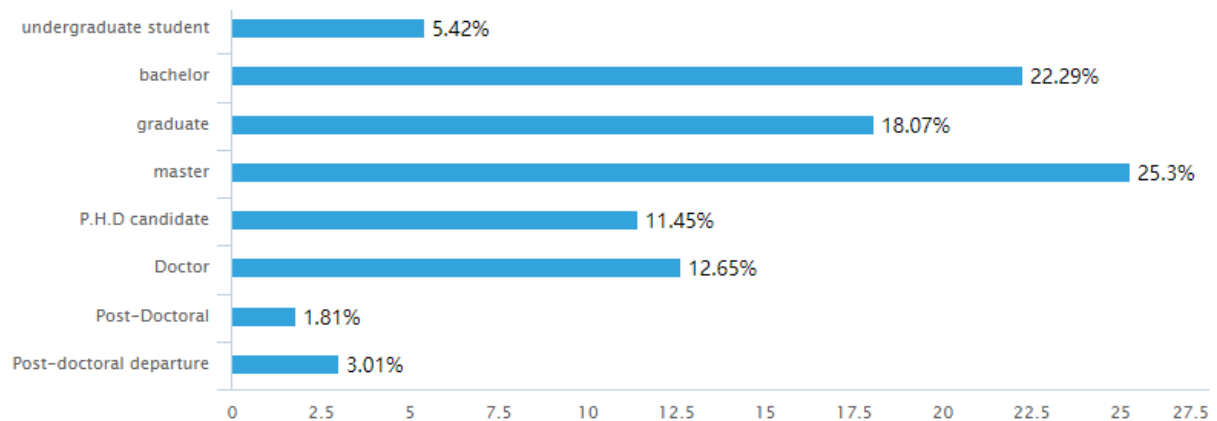


Figure 2 Percentage of researcher categories among 160 inquiries

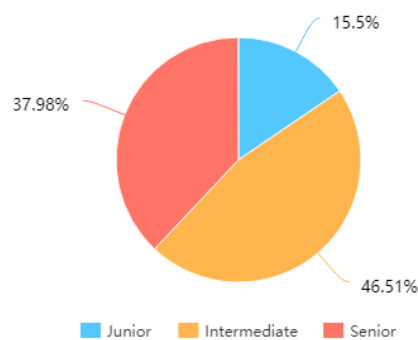


Figure 3 Title percentages of interviewees from China mainland

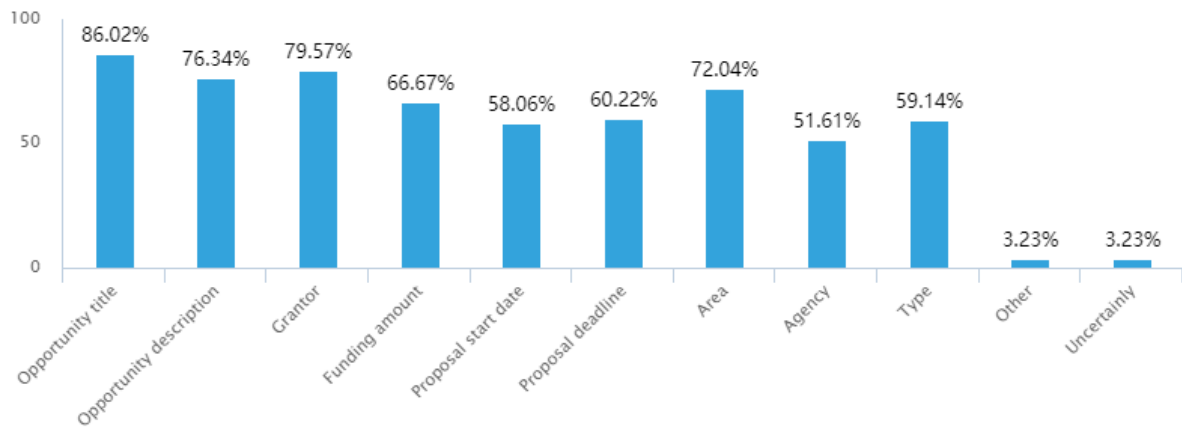


Figure 4 Attribute needs for forecast projects

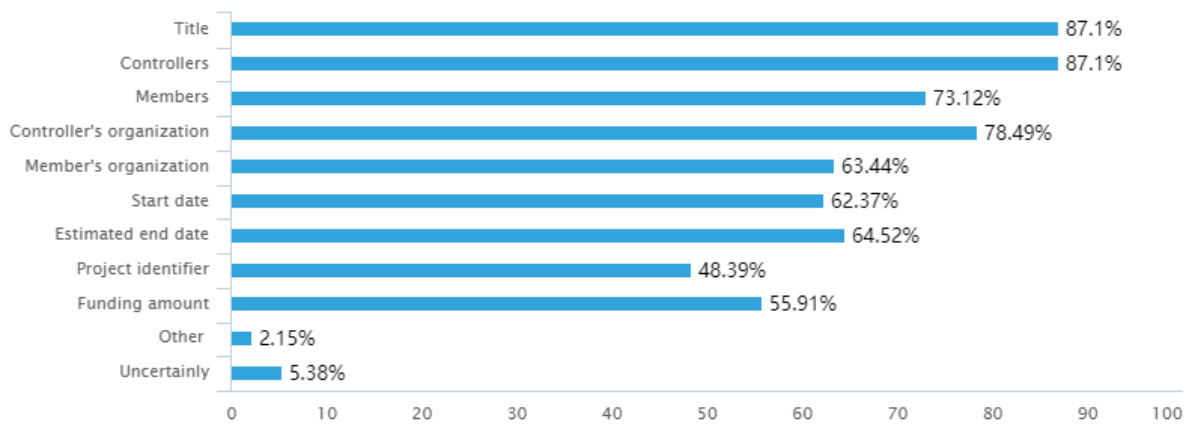


Figure 5 Attribute needs for granted projects

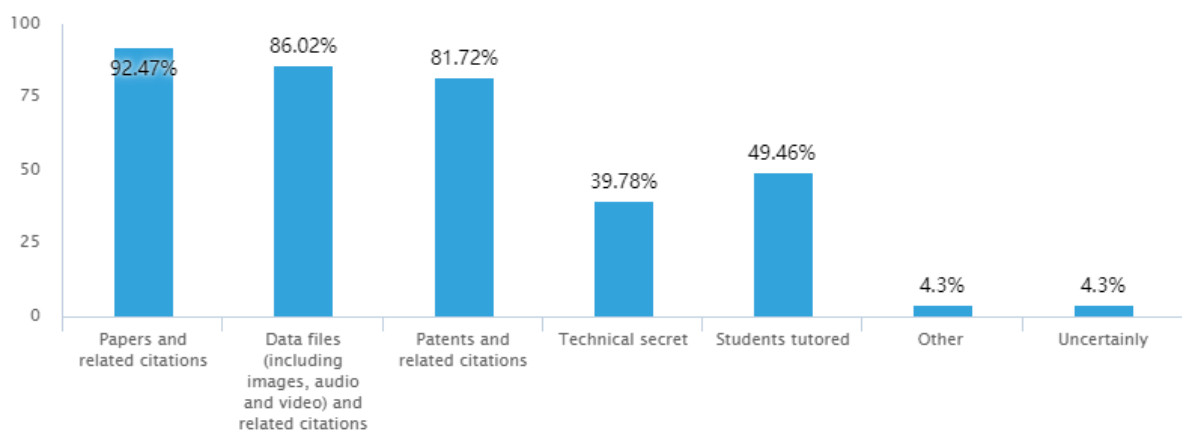


Figure 6 Attribute needs for finished project

For the question "What do you think the research projects database can be used for?" The top three answers were all related to discovering of research problems, research opportunities and

construction of application strategies with popularity votes of 85.95%, 72.04%, and 65.59% respectively. As shown in table 2, the utilization of research outputs and cooperation seeking were also more concerned.

Table 2 expectation popularity of the usage of research projects

Usage	Vote	Percentage%
Discovering research opportunities	67	72.04
Discovering research problems	79	84.95
Visiting	36	38.71
Construction of application strategies	61	65.59
Capital Performance Analysis	29	31.18
Analysis of Financial Aid	30	32.26
Research output utilization	51	54.84
Seeking cooperation	50	53.76
Human resource recruitment	28	30.11
Others	1	1.08
Total	93	100

For the inquiry "Some people think that it is better to have a database to integrate opportunity information, granting information, research outputs and dataset of research projects. To what extent do you agree at present? [from 9 (most agree) to 0 (not agree)]", there were 45 votes for agreement, 12 votes for more agreement and 39 votes for most agreement. This means that most researchers in this survey would like research projects to be integrated with its research outputs.

For the inquiry "When you read the reference, you want to know the following information about the project of the reference at the same time. To what extent do you agree at present? [from 9 (most agree) to 0 (not agree)]", only a few of 93 researchers voted for not agreement or uncertainty, as shown in table 3. From this, we can infer that researchers would like to find

research problems or application opportunities with the help of the information on related projects as well as information concerning the usage of the research project.

Table 3 Agreement to activate the link out from one research publication with related projects

Information type	0 (not agree)	3	5	7	9 (most agree)	Uncertainly
Other research output	3(3.23%)	0(0%)	40(43.01%)	23(24.73%)	24(25.81%)	3(3.23%)
Application materials	4(4.3%)	5(5.38%)	32(34.41%)	21(22.58%)	28(30.11%)	3(3.23%)
Funding information	2(2.15%)	4(4.3%)	38(40.86%)	23(24.73%)	22(23.66%)	4(4.3%)

Metadata of research project: an analysis sheet

The object of this step is to delete some redundancy attributes which may be less important for sharing and set some vocabularies as candidates. First, we set vocabularies from Grant.Gov and Research Professional as the basement controllers of authority. Second, All vocabularies referring attributes of research projects or related outputs were listed by categories. For forecasting information, there were nine types of categories, such as agency, contact, content, date, .etc. For granting information, there were seven types, such as content, contributors, date, funding money, identity, status. There only two categories for finished research projects attribute vocabularies. Third, as mentioned before, most researchers who replied the survey expected to be inspired by the project information and the research output related to projects when reading a reference, finding research opportunities and research problems. As shown in table 4, most of the vocabulary candidates were less important for the important usage of research projects, such as grant officer name, key contact name. Others were specific attributes defined by some organizations, such as CFDA number(s) and QFIS project code. Identity was not very concerned by the interviewees, but it can be used as an identification key value to find a project.

Table 4 Vocabularies for research projects

Research projects state	Information type	Vocabulary	Vocabulary candidate
forecast	contact	agency name email phone	grant officer name key contact name
	content	description opportunity category explanation opportunity category opportunity title program name	category explanation category of funding activity document type document version
	dates	proposal deadline proposal deadline time zone	decision date fiscal year forecast date last update date
	eligibility	application instructors eligible applicants information on eligibility research project criteria	
	funding scale	award ceiling award floor cost sharing estimated total program funding expected number of awards	
	grantor	grant maker link to additional information program guidelines	CFDA number(s) fund source

	identities	opportunity number	
	status	archived closed forecast posted	
Granted	content	abstract title subjects	acronym description Keywords type
	contributors	country organizations of principal investigator principal investigator(PI)	administrator co-investigator researcher promoter
	date	effective start date end date	
	funding money		funding money
	identity	id	QFIS project code
	status	status	
finished	General	abstract	archive date description keywords
	Output	articles conclusion reports data sets impact patents	Activities final reports midterm reports papers secrets services

Research projects as nodes

It has been a long time that research projects earned an embarrassing role in research workplace. On one hand, from this survey, over 40% researchers of 93 interviewees mentioned above knew the fact that research project is the important indicator for both year-end assessment and promotion, over 54% of them voted the point of view that granted chance is very low for normal researchers and only 44.09% of them reported the fact of licensed databases of research projects. But for the question "How do you choose your thesis or project?" only 37.65% of them preferred to find research problems from callings for research projects opportunities, details are shown in table 5. On the other hand, even some products have designed research projects as resource nodes, some research customers would probably not prefer it. For example, there are four types of site navigates on the web pages of Pure repositories, only two of them set projects as a same level node as research output or publications.

The thinking of metadata for research projects may encourage the actions of setting the research projects as information resource nodes. At this point, we found that the resource node webs of Research Professional could be used as a nodes web basement and modify it by creating links between research projects, datasets, publications, and other related projects. For example, as shown in figure 7, project 1 is related to project 3, project 3 is related to project 4, all these projects have links to publications and datasets.

Table 5 The percentage of methods to find research topics

Idea source	vote	Percentage%
Inspired by references	70	75.27
Problems from workplaces	70	75.27
Calling of journals or conferences	39	41.94
Calling of projects proposals	35	37.63
Calling from enterprises	3	3.23
Projects from tutors	25	26.88
Projects from leaderships	13	13.98
Projects from colleagues outside workplaces	18	19.35
Projects from colleagues of workplaces	14	15.05
Other	1	1.08
No experiences of projects proposal	3	3.23
No experiences of article writing	1	1.08

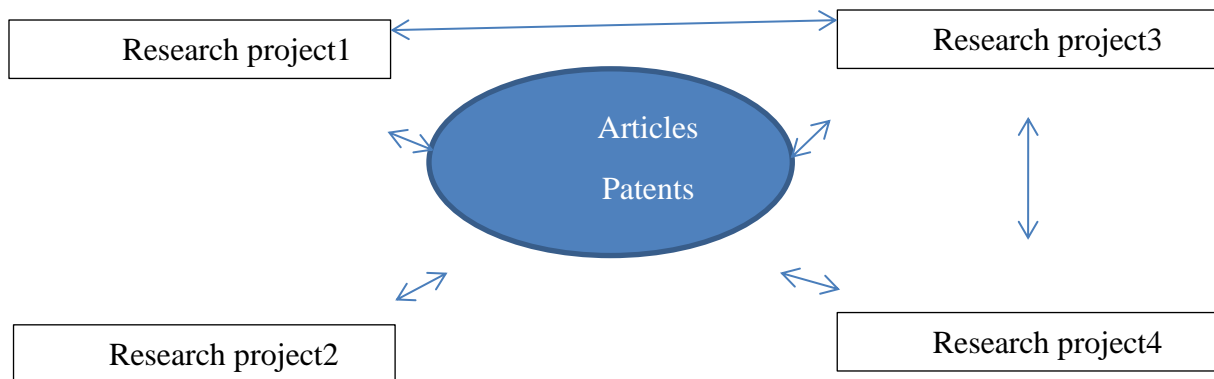


Figure 7 The research projects in the node web

Conclusions and Discussions

This article proposed an analysis metadata vocabulary sheet for research projects by using the method of investigation. The vocabularies were classified to 16 categories referring to 3 status of research projects. The investigation focused on research project attribute vocabularies from 12 databases or repositories concerning research projects and acknowledgment references of research projects. The innovation of this article is the survey aiming to collect the references needs, granting opportunity circumstances and the usage of research projects databases, which created an opportunity for metadata vocabulary analysis with the help of the conclusion of the survey. Once the research project metadata comes out, the resource nodes webs can be generated based on the sharing and harvesting of research projects information.

Acknowledgments

I would like to thank East China University of Science and Technology for supporting me do my research work as a visiting scholar at School of Information Science of University of Illinois at Urbana-Champaign, which should be grated for their course about metadata too. I am also grateful to Jiangsu Zhongyang Data Technology Co., Ltd. for supporting me to attend WLIC 2019 at Athen. Lei Wang from Jiangsu Zhongyang Data Technology Co., Ltd., Qingfang He and her group members from Tongji University Library should be included in the acknowledgments list for inspiring me and Weihong Luo getting into the field of metadata of research projects.

References

1. Priscilla Caplan (2003). Metadata Fundamentals for All Librarians. ALA Editions, ProQuest Ebook Central. <http://ebookcentral.proquest.com/lib/uiuc/detail.action?docID=3001579>.

2. Andres Melgar, Linder Corro. (2016). A Framework for Organizational Memory Management of Research Projects in Institutions of Higher Education[C]. 2016 11Th Iberian Conference on Information Systems and Technologies (CISTI) Series: Iberian Conference on Information Systems and Technologies.
3. Antonella Fresa, Borje Justrell, Claudio Prandoni. (2015). Digital Curation and Quality Standards for Memory Institutions: PREFORMA Research Project[J]. Archival Science, 15 (2): SI, p. 191-216.
4. Stepan Oana Maria, Lungu Mircea-Alexandru. (2015). Specific Research Projects Grant of Semantically[C]. 2nd International Multidisciplinary Scientific Conference on Social Sciences and Arts (SGEM 2015), Albena, Bulgaria, Aug 26-Sep 01
5. Constanze Curdt, Dirk Hoffmeister.(2015). Research Data Management Services for a Multidisciplinary, Collaborative Research Project Design and Implementation of the Tr32Db Project Database[J]. Program-Electronic Library and Information Systems, 49(4): SI, p, 494-512
6. Weisen Guo, Steven B Kraines.(2011). Discovering Relationship Associations from the Literature Related to Research Projects in Sustainability Science Using Ontology and Inference[C]. International Conference on Knowledge Discovery and Information Retrieval, Univ Paris Est Creteil, Paris, France, OCT 26-29
7. S. Shreeves, J. Riley, L. Milewicz. (2006). Moving Towards Shareable Metadata[J]. First Monday 8 (7), <https://doi.org/10.5210/fm.v11i8.1386>
8. John L. McCarthy. (1982). Metadata Management for Large Statistical Databases[c].The Proceedings of the Eighth International Conference on Very Large Data Bases, Mexico City, Mexico, Sep 8-10
9. Becker R. A. , Chambers J. M.. Design and Implementation of the 'S' System for Interactive Data Analysis[C]. Proceedings of I.E.E.E. Compsac78, p. 626-629.
10. Teitel R. F. (1977). Relational Database Models and Social Science Computing[C]. Proceedings of Computer Science and Statistics: Tenth Annual Symposium on the Interface, Gaithersburg, MD: National Bureau of Standards, p. 165-177.
11. L. Mark, N. Roussopoulos. (1986). Metadata Management[J]. Computer, 19(12), p. 26-36
12. PMD Gray, GE Storrs, JBH Duboulay. (1988). Knowledge Representations for Database Metadata[J]. Artificial Intelligence Review, 2(1): 3-29.
13. A. Alzobaidie, JB Grimson. (1988). Use of Metadata to Drive the Interaction Between Database and Expert Systems[J]. Information and Software Technology, 30(8), p. 484-496.
14. M. Siegel, SE. Madnick. (1991). A Metadata Approach to Resolving Semantic Conflicts[C]. 17th International Conference on Very Large Data Bases, Location: Barcelona, Spain, Date: Sep 03-06.
15. YS Joung, JH Lim, SJ Hyun. (2001). A Metadata Repository System For an Efficient Description of Visual Multimedia Documents[J]. Concurrent Engineering-Research and Applications, 9(2), p. 93-104.

16. AKMS Islam, M. Piasecki. (2006), A Generic Metadata Description For Hydrodynamic Model Data[J]. Journal of Hydroinformatics, 8(2), p. 141-148.
17. Andras Hajdu, Charalambos Giamas, Nicholas Vretos. (2007). Metadata Description of Thermal Videos for Rescue Operations[C]. International Symposium on Signals, Circuits and Systems, Location: Iasi, Romania, Date: JUL 12-13.
18. Hong Yao, Jinlai Xu, Zhongwen Luo. (2016). MEMoMR: Accelerate MapReduce Via Reuse of Intermediate Results[J]. Concurrency and Computation-Practice & Experience, 28 (14), p. 3814-3829
19. Athanasios D. Styliadis, Ipek I. Akbaylar, Papadopoulou, Despoina A.(2009). Metadata-based Heritage Sites Modeling With e-Learning Functionality[J]. Journal of Cultural Heritage, 10(2), p. 296-312.
20. Petr Kremen, Martin Necasky,(2019). Improving Discoverability of Open Government Data With Rich Metadata Descriptions Using Semantic Government Vocabulary[J]. Journal Of Web Semantics, 55, p. 1-20.
21. Xiaobin Li, Peijie Zhuang, Chao Yin. (2019). A Metadata Based Manufacturing Resource Ontology Modeling in Cloud Manufacturing Systems[J]. Journal Of Ambient Intelligence And Humanized Computing, 10(3-special: SI), p. 1039-1047
22. Grace Therrell. (2019). More Product, More Process: Metadata in Digital Image Collections[J]. Digital Library Perspectives, 35(1), p. 2-14.
23. Qingfang He. (2017). Quanqiu Keyan Xiangmu Shujuku Yingyong Yu Xueke Jianshe Youxiao Lujing Zhi Tansuo[R].