

Geospatial Collections in Institutional Repositories (IRs): a survey in Map/GIS Libraries

Ifigenia Vardakosta*

Laboratory on Digital Libraries and Electronic Publishing
Department of Archive, Library and Museum Sciences
Faculty of Information Science and Informatics
Ionian University
Ioannou Theotoki 72, Corfu 49100, Greece
ifigenia@ionio.gr

Harokopio University
Library and Information Centre
El.Venizelou 70, 17671 Kallithea
ifigenia@hua.gr
tel: +302109549170

Sarantos Kapidakis

Laboratory on Digital Libraries and Electronic Publishing
Department of Archive, Library and Museum Sciences
Faculty of Information Science and Informatics
Ionian University
Ioannou Theotoki 72, Corfu 49100, Greece
sarantos@ionio.gr

Abstract

Changes in scholarly publishing along with the rapid developments in technology are affecting user expectations and forcing academic libraries to significant changes in providing access to geospatial data and develop new services. Institutional Repositories (IRs) are considered effective tools for showcasing an organization's scientific research by making it publicly available.

The aim of this study was to determine the awareness of the development of locally produced geospatial data in an institutional repository, among the Map/GIS libraries of the western world. Anonymous survey of 56 questions, divided in 6 sections and covering several topics, was sent to Map/GIS libraries, members of Map and Libraries' Organizations of western world. The survey took place on July and October 2015 after two phases of pilot distribution. Out of 382 questionnaires we sent, a total of 80 were finally adequate to be used. This paper will present those results related with the collection of geospatial data in the IR (36 responses) produced by the members of library's academic community.

The responses showed a statistically significant trend to the development of a locally produced geospatial data collection in the IR, although the annual budget is considered by the Map/GIS Librarians as “not at all satisfying” while in majority respondents indicate the use of policies for the locally produced geospatial data collection in the repository. As the three most important issues for this collection, respondents indicate: 1. data and metadata 2. policies and 3. finance.

The development of a geospatial data collection is rather complex particularly with respect to issues related to organizational challenges and use practices, therefore, Map/GIS Librarians seem to be convinced that the formulation of policies, the open access to the data will be the critical points to accomplish a beneficial collection of the locally produced geospatial data in the IR.

1. Introduction

The last decade IRs has been a major part of library's activities and the main promotion source of the intellectual production for the faculty and researchers of the parent institution. Repositories facilitate the dissemination, access, reuse and preservation of the work arising from the scientific and academic activity of their staff (Serrano-Vicentre et al. 2016). Research data-in which geospatial data are included as well- are important output of the scholarly research process across all disciplines. Data's value increases as they are aggregated into collections and as they become more available for reuse to address new research questions (Kim 2013). Additionally, without knowing what data already exists on campus, it is possible for multiple departments to create the same data sets. Libraries can serve as a geospatial data repository thus providing a centralized location to access all geospatial data on campus (Howser and Callahan 2004). Moreover research funders have increasingly begun to mandate the creation of research data management plans and the deposit of research data in recognized data centers and IRs (Koltay 2016). For instance, funding agencies such as the National Institutes of Health are beginning to require the deposit of publications derived from the research that they sponsor into open access repositories. Similarly, some funding agencies such as the National Science Foundation are moving toward requiring that grant proposals include data management plans that address preservation and open access to the data that is generated by their sponsored projects (Witt 2009).

Integrating research data collections is however a complex task for any library since there are challenges that need to be addressed and resolved (e.g. copyright, data licensing, data security, data privacy) (Patel 2016). The need for data sharing has been recently more emphasized as data-intensive science continues to grow and the open data movement gain traction around the world (Kim 2013).

As academic libraries must continually adapt to the changing landscape of higher education (Cerbo 2012), the present study seeks to reveal whether Map/GIS Libraries have expanded their services in response to the rise of the data driven research. Therefore, the methodology used in the research is presented thoroughly, while descriptive results will be given in detail. After that the statistical validation of the results will be presented while the conclusions and the further work will close the paper.

2. Literature Review

Changes in scholarly publication and a rapid development in technology are forcing libraries to develop new resources and services areas. As Jantz and Wilson (2008) point out the literature according the connection of an IR development with the open access movement is expanding rapidly. Sponsors from the European Commission, the U.S. Government, and the Australian Government and other players with the data community launched the Research Data Alliance (RDA)¹. This initiative aims to facilitate the global research data sharing and exchange by the harmonization of data standards and practices. The Open Data Institute (ODI)² was launched in the UK with the objective to promote new business and culture around open data by creating economic, environmental, and social value and by promoting standards. Confederation of Open Access Repositories (COAR)³ aims to position universities, research institutions and their libraries as the foundation for a new scholarly communication system. In that system, where openness is becoming the default, it is imperative that institutions take on greater responsibility for the collection, management and sharing of this valuable content. COAR launches several activities for 2016-2018, including among others, the share of the collection policies for research data management in repositories. European Commission also sustains HORIZON 2020⁴, the biggest EU Research and Innovation programme while RECODE⁵ the project funded by the European Union will provide a space for European stakeholders, interested in open access to research data and work together in order to provide common solutions for these issues. The INSPIRE Directive⁶ aims to create a European Union spatial data infrastructure for the purposes of EU environmental policies and policies or activities which may have an impact on the environment. “*Institutional Repositories*” (57.1%), “*collections*” (81.01%) and “*open access*” (23.8%) were among the themes listed as the most common priority areas in Saunders study of academic libraries strategic plans (2015). Newton, Miller & Brack (2011) point out that the collection practices for research data set change in the library when the target collections are produced locally while Morris (2006) highlights that libraries focuses discussion on commercially published datasets. Among the actions for making an IR more successful and lasting is the provision of clear policies as researchers suggest (Marsh 2015; Priti 2011). Kenyon, Godfrey and Eckwright (2012) describing their experience in developing the “*Numeric and Spatial Data Information Data Engine*” (INSIDE Idaho) include policies among issues that academic libraries must develop in order to manage and curate research data successfully.

The current research attempts to fill the gap in the Map/GIS Libraries literature determining whether the latest approaches regarding policies and open access have influenced the collection of locally produced geospatial data.

¹ <https://rd-alliance.org>

² <http://theodi.org>

³ <https://www.coar-repositories.org/>

⁴ <https://ec.europa.eu/programmes/horizon2020/>

⁵ <http://recodeproject.eu/>

⁶ <http://inspire.ec.europa.eu/>

3. Methodology of the Research

4.1. Sample

The survey addressed to the Map/GIS Librarians whose libraries were members of Map/GIS Libraries Associations. As Map/GIS Libraries Associations lead and inspire the professional group of Map/GIS Librarians and others interested in geographic information, we considered their members as the ideal sample for covering the needs of this research.

So, the survey was addressed to the members of “Map and Geospatial Information Round Table of the American Library Association” (MAGIRT)⁷, “Western Association of Map Libraries” (WAML)⁸, “Association of Canadian Map Libraries and Archives” (ACMLA)⁹ “Australian and New Zealand Map Society” (ANZMapS)¹⁰ and “Map and Geoinformation Curators Group” (MAGIC)¹¹. Although ARL-GIS Literacy Project no longer exists, its significant contribution to the spread of digital spatial data could not be ignored. That’s why we included those libraries-members of the project to the research.

The next step for finalizing our sample was to merge all lists to one, since in many cases (mainly in U.S. and Canada) a library could be a member in more than one Associations (e.g. Pennsylvania State University is a member of both MAGIRT and WAML Association and also a member of ARL-GIS Literacy Project). The above action has resulted in the creation of a final spreadsheet in which 382 libraries appeared only once, along with the e-mail of the personnel responsible for Map/GIS collection.

4.2. Questionnaire

To access the group of librarians needed for the research the electronic questionnaire ability that Google Forms hosts was used after modification in such a way that could cover survey’s special needs. The customization ability along with other technical reasons (e.g. it provides a unique link to the survey, the results can be saved directly in a spreadsheet, allows the researcher to control the unique participation and to be the only one who has access to the data, ensuring their protection and reliability) were the main reasons for the use of this tool.

The questionnaire was divided in 6 sections covering various topics: 1) Demographics: for the respondent, the organization and the library (14 questions), 2) Geospatial Collection Development (12 questions), 3) Geospatial collections in Institutional Repository (11 questions), 4) GIS services (6 questions), 5) Policies (5 questions), 6) Opinions (8 questions). Detailed instructions were given considering the completeness and the purpose of each section. Most of the questions were closed-ended and some permitted multiple answers. According to their given answers, participants could direct from one section to another.

The survey was anonymous; therefore none of the survey questions asked participants to provide any details about themselves or their every day job in their working environment.

⁷ <http://www.ala.org/magirt>

⁸ www.waml.org

⁹ www.acmla-acacc.ca

¹⁰ www.anzmaps.org

¹¹ <http://cartography.web.auth.gr/ICA-Heritage/MAGIC/index.html>

The survey took place on July and October 2015 after two phases of pilot distribution. Out of 382 messages we sent with a link to the questionnaire, a total of 80 were finally adequate to be used. This paper will present those results related with the collection of geospatial data in the IR (36 responses) produced by the members of library's academic community.

4.3. Research Questions

The aim of this study was to determine the awareness of the development of locally produced geospatial data in an institutional repository, among the Map/GIS libraries of the western world.

More specifically the research questions that this research intends to answer are as follows:

RQ1: Do Map/GIS librarians collect the local produced geospatial data in their IRs?

RQ2: Are Map/GIS librarians using certain policies for developing IR's geospatial collections?

RQ3: Is open access a distribution choice?

5. Research Results

Research results are reported here by theme, drawing together the findings.

5.1. Demographics

The purpose of this section is to reveal demographic information for the sample and identify their working environment in terms of organization and library information.

Out of 36 responses to the questionnaire 52.8% (n=19) were from USA, 33.3% (n=12) were from Europe, 8.4% (n=3) were from Australia and 5.5% (n= 2) responses were from Canada.

As the composition of Map and GIS Association Lists reflects a variety of library types, our sample includes a majority of academic libraries (80.5%, n=29), followed by National Libraries (13.9%, n=5) and State Libraries (5.6%, n=2).

Respondents were in majority "*librarians*" (36%, n=13), "*GIS librarians*" (31%, n=11), "*Map librarians*" (11%, n=4), "*Library Directors*" (11%, n=4) and "*GIS specialists*" (11%, n=4).

The working experience declared by the majority of individuals was "*<5 years*" (40%, n=14), "*>17 years*" (26%, n=9), while "*12-16 years*" and "*6-11 years*" (14%, n=11) was declared by 17%, n=6 accordingly.

In the next group of questions regarding their *organizations*, respondents in vast majority (72%, n=26) indicated that their institutions "*sustained a department related to geospatial information and GIS*" while 28% (n=10) responded negatively to the question. Asked about the *population of their institution*, respondents stated in majority that they are working in an institution which its population exceeds "*10001+*" (67%, n=24), while 14% (n=5) declared population "*-2000*" for their institution. Population average "*2001-5000*" indicated by 11% (n=4) of the respondents and finally population average "*5001-10000*" indicated by only 8% (n=3). The *institutional annual budget* is considered for the 63.9% (n=23) of the respondents as "*moderately satisfying*" and for the 22.2% (n=8) as "*slightly satisfying*", while for

13.9% (n=5) is and “*not at all satisfying*”. None of the participants consider annual budget as “very” or “extremely satisfying”.

The vast majority of the participants (86%, n=31) are working in *libraries* with more “21+” staff, while 8% (n=3) declare “0-5”. Only one of the participants (3%, n=1) declared that his/her library has “6-10” and “16-20” number of personnel accordingly. Respondents stated in majority (61%, n=16) that their library sustains “10001+” registered users, while “-2500” number of users stated by 15% (n=4), “5001-8000” number of users stated by 12% (n=3) of the participants. Finally, 8% (n=2) of the respondents stated “8001-10000” as number of registered library users and only 4% (n=14) of the respondents declared the number of their library users as “2501-5000”. Research respondents reported in majority that in their opinion the library’s annual budget is “*moderately satisfying*” (38.9%, n=14). As “*satisfying*” the annual library budget is indicated by the 25% (n=9), of the respondents while the 22.2% of them (n=8) believes that it «*is not at all satisfying*». Finally, only the 13.9% (n=5) believes that the library’s annual budget is “*very satisfying*”.

As respondents indicate the *IR in their institution developed* by the “*Library*” (61%, n=22), by “*the Library in cooperation with IT*” (33%, n=12), and by the “*IT*” and the “*Institutional Office for Research*” (3%, n=1) accordingly.

In Table 1, responses in the “Demographics-General Questions” section are reported in majority.

QUESTIONS	RESPONSES	FREQUENCIES	PERCENTAGE
RESPONDENTS			
Country	USA	19	52.8%
Institution type	Academic		
	Libraries	29	80.5%
Participant Specialty	Librarians	13	36%
Working experience	<5 years	14	40%
ORGANIZATIONS			
Department related to GIS	Yes	26	72%
Organization’s population	10001+	24	67%
Organization’s annual budget	moderately satisfying	23	63.9%
LIBRARY			
Number of Library’s personnel	21+	31	86%
Library’s population	10001+	16	61%
Library’s annual budget	moderately satisfying	14	38.9%
IR Development	Library	22	61%

Table 1: Demographic details and general questions (in majority)

5.2. Locally produced geospatial data collection in the IRs

The collection of the local produced geospatial data in the IR has been developed, as respondents indicate in majority, “*by the library*” (44.4%, n=16), based on certain

policies (61.1%, n=22), the number of librarians involved in geospatial collection are “*1-2*” (60%, n=22), while the same number i.e. “*1-2*” declared as the one that have been trained/educated for geospatial data (56%, n=20). The geospatial collection in IRs consists of a variety of geospatial data mainly “*digital maps*” (83.3%, n=), in “*images*” format (e.g. tiff, jpeg, png) while one of the main categories that are collected is “*topography*” (83.3%). The participants in the survey consider the development of the collection as “*good*” (36%, n=13), while they believe that the *annual budget* for the locally produced geospatial data collection in the repository is “*not at all satisfying*” (46%, n=15).

In our question for pointing out the 3 most important issues regarding the geospatial collection at this period of time in their library respondents declared 1. data and metadata (55.5%, n=20), 2. policies (52.7%, n=19), and 3. finance (50%, n=18). In Table 2, responses in the “IRs” section are reported in majority.

QUESTIONS	RESPONSES	FREQUENCIES	PERCENTAGES
IR geospatial collection development	By the Library	16	44.4%
number of librarians involved in geospatial collection	1--2	22	60%
number of librarians trained/educated for geospatial data	1--2	20	56%
collection development	Good	13	36%
policies use	Yes	22	61.1%
annual budget	not at all satisfying	15	46%
3 most important issues	1. data and metadata	20	55.5%
	2. policies	19	52.7%
	3. finance	18	50%

Table 2: Respondent's answers regarding geospatial collection in an IR in majority

5.3. Open Access in IRs

Participants declared that in their IRs, the content of the locally produced geospatial data is “*accessible and reusable by everyone*” (61.1%, n=22), while the easiness for a library to adopt open access in a repository collection is rated as “*neutral*” (36.1%, n=13). Participants in the survey in majority “*agree*” and “*strongly agree*” (41.7%, n=15 and 44.4%, n=16 accordingly) that “*libraries should undertake initiatives, adopt open access and promote the local produced geospatial data*” (Fig.1) since they consider open access as a “*very important*” (41.75%, n=15) action towards a geospatial collection development.

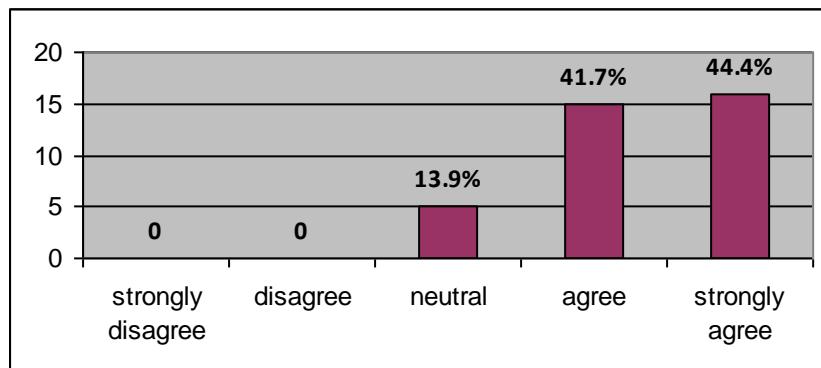


Fig.1: Libraries should undertake initiatives, adopt open access and promote the local produced geospatial data

6. Statistical analysis

The statistical analysis was carried out using the SPSS version 24 software package.

6.1. Test of independence

Additionally we applied a chi square test as it is a test that can be used to evaluate whether or not frequencies which have been empirically obtained differ significantly from those which would be expected under a certain set of theoretical assumptions (Blalock 1979: 279). The test was applied for the variables “how important is open access in the development of a geospatial collection” and “how important are policies for the geospatial collection development”. The test revealed p-value = $0,084 < 0,10$ consequently variables are dependent in statistical significance $\alpha=10\%$ (Table 5).

Chi square test	
X^2	p-value
2.991	0.084

Table 3: Chi-square test for “open access” and “geospatial collection”

6.2. Normal Distribution

The low sample number led us to apply a Kolmogorov-Smirnov Test. The following couple of hypothesis formulated for checking normal distribution:

H_0 : The variable follows normal distribution (null hypothesis).

H_1 : The variable does not follow normal distribution (alternative hypothesis).

The data in Tables 7, reveal that p-value in all variables is equal with $0,000 < 0,005$, a fact that leads to reject the null hypothesis, and accept the alternative hypothesis i.e. the variables does not follow the normal distribution.

		IR's policies	IR's GC level	IR's GC budget	IR's GC OA
N		36	36	36	36
Normal Parameters ^{a,b}	Mean	1.58	2.61	2.03	2.92
	Std.	.806	1.271	1.028	.996
	Deviation				
Most Extreme Differences	Absolute	.376	.231	.258	.189
	Positive	.376	.203	.258	.189
	Negative	-.235	-.231	-.189	-.172
Test Statistic		.376	.231	.258	.189
p-value		.000 ^c	.000 ^c	.000 ^c	.002 ^c

a. Test distribution is Normal.

b. Calculated from data.

c. Lilliefors Significance Correction.

Table 4: One-Sample Kolmogorov-Smirnov Test

7. Limitations of the research

It is considered that two key parameters contributed to the low participation rate in the survey: The first one was the questionnaires' length which was prohibited although all the technical facilities were employed allowing respondents to participate in the sections of their interest preventing any unnecessary delay. The second factor that didn't work in favour of the research was the time of the questionnaire's dissemination as summer is a holiday period and most of the recipients were not at their workplaces.

8. Discussion

As findings revealed large Map/GIS libraries recognizing their role in providing significant preservation and dissemination services, tend to make efforts towards the development of IR collections for their local geospatial data. Map/GIS librarians are fully equipped to assist faculty and researchers to organize their geospatial research data. The research determines that the majority of the existing collections are based on certain policies. This fact conflicts with the results of previous surveys for the use of policies in geospatial collections (Vardakosta and Kapidakis 2016). Despite the differences in methodology it seems that digital environments' requirements along with the peculiarities of geospatial data require their establishment.

Most of the Map/GIS librarians are not satisfied with the development level of the IR collection nor its annual funding. This point of view 'causes no surprise since libraries worldwide have been affected by the global economic depression (Guarria and Wang 2011). Results came to confirm Lowery's (2016) survey which was conducted to

provide information on the consequences to map department budgets in academic libraries from recent economic depression. Among the findings was the intention of the map librarians who participated in the study to seek for alternative funding approaches for their collections, such as crowd funding.

Their answers for the most important issues regarding the collection in the IR, focus on data and metadata while open access is considered a well known procedure and quite important for IR's development. These findings from Map/GIS librarians conflict Saunders' (2015) research in which data management and open access included in fewer than half of the plans reviewed despite the fact they were both highlighted as top trends by ACRL¹². Additionally, Map/GIS librarians recognize the importance of data and metadata and they rank those two issues as priorities. As data form the collection, metadata are crucial for the presentation of data to the wider audience and contribute that way to its appropriate use (Forward, Leahey and Trimble 2015).

9. Conclusions

This research brings together the Map/GIS Librarians answers regarding their awareness in the development of a local produced geospatial data collection in an IR. The evidence indicate that the development of a collection of locally produced geospatial data is a trend of the large Map/GIS libraries reflecting their adjustment to current scholarly environment and their users information needs. Despite the lack of finance, Map/GIS librarians undertake initiatives, establish policies, adopt open access and promote data sharing between creators and users of spatial data through a centralized mechanism such as the IR. Policies and open access are issues that Map/GIS librarians apply for the beneficial implementation and use of the collection. Dealing with the continuously growing geospatial information Map/GIS librarians respond dynamically. The establishment of a locally produced geospatial data collection serves a twofold purpose. At first it responds to the faculty needs for preservation, process, wider acceptance of their work, and to their obligation to the programme that funded their work. Secondly, it consists a low budget chance for Map/GIS libraries to provide their users with a new collection and related services. Map/GIS Librarians seem to be convinced that the formulation of policies and the open access to the data will be the critical points to accomplish a beneficial collection of the locally produced geospatial data in the IR.

Acknowledgements

Authors wish to thank all the participants in the pilot distribution of the questionnaire, the Map/GIS Librarians that distributed the questionnaire in the several topic mail-lists and also all those professionals that kindly responded. Additionally, authors wish to thank the Map/GIS Librarians that provided great help with those Associations' member lists that were not in public sight. The research was not conducted under any financial support.

¹² Association of College and Research Libraries (<http://www.ala.org/acrl/>)

BIBLIOGRAPHY

- Cerbo, M.A. (2012). The Academic Library Online: Is the Future of Academic Libraries a Virtual Reality? *Technical Services Quarterly*, 29 (3), 181-192, DOI:10.1080/07317131.2012.681278
- Forward, E., Leahey, A. & Trimble, L. (2015). Shared Geospatial Metadata Repository for Ontario University Libraries: collaborative Approaches. *New Review of Academic Librarianship*, 21 (2), 170-184. DOI: 10.1080/13614533.2015.1022662
- Guarria, C.I. and Wang, Z. (2011). The economic crisis and its effect on libraries. *New Library World*, 122 (5-6), pp.199-214
- Howser, M. & Callahan, J. (2004). Beyond locating data: academic libraries role in providing GIS services, 12 p. Retrieved from <http://proceedings.esri.com/library/userconf/educ04/papers/pap5127.pdf>
- Kenyon, J., Godfrey, B. & Eckwright, G.Z. (2012). Geospatial Data Curation at the University of Idaho. *Journal of Web Librarianship*, 6 (4), 251-262, DOI:10.1080/19322909.2012.729983
- Jantz, R.C. & Wilson, M. C. (2008). Institutional Repositories: Faculty Deposits, Marketing, and the Reform of Scholarly Communication. *The Journal of Academic Librarianship*, 34(3), pp.186-195
- Kim, J. (2013). Data sharing and its implications for academic libraries. *New Library World*, 114 (11/12), pp. 494-506
- Koltay, T. (2016). Facing the Challenge of Data-Intensive Research: Research Data Services and Data Literacy in Academic Libraries. *Innovation in Libraries and Information Services*.
DOI: <http://dx.doi.org/10.1108/S0732-067120160000035008>
- Lowery, R. (2016). Funding in a Time of Recession: Map Collection Revenues. *Journal of Map & Geography Libraries*, 12(2), 202-215, DOI:10.1080/15420353.2015.1105900
- Marsh, R.M. (2015). The role of institutional repositories in developing the communication of scholarly research. *OCLC Systems & Services: International digital library perspectives*, 31 (4), pp.163 - 195
- National Institutes of Health. (2008). Revised policy on enhancing public access to archived publications resulting from NIH-funded research. <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-08-033.html>
- Newton, M.P., Miller, C.C. & Stowell Bracke, M. (2011). Librarian Roles in Institutional Repository Data Set Collecting: Outcomes of a Research Library Task Force. *Collection Management*, 36, 53-67

Patel, D. (2016). Research data management: a conceptual framework. *Library Review* 65 (4/5), pp.226 – 241. DOI <http://dx.doi.org/10.1108/LR-01-2016-0001>

Priti, J. (2011). New trends and future applications/directions of institutional repositories in academic institutions. *Library Review*, 60 (2), pp.

Saunders, L. (2015). Academic Libraries' Strategic Plan: Top Trends and Under-recognized Areas. *The Journal of Academic Librarianship*, 41 (3), 285-291

Serrano-Vicente, R., Melero, R. & Abada, E. (2016). Open Access Awareness and Perceptions in an Institutional Landscape. *The Journal of Academic Librarianship* 42 (5), p.595-603. <http://dx.doi.org/10.1016/j.acalib.2016.07.002>

Vardakosta, I. & Kapidakis, S. (2016a). Geospatial Data Collection Policies, Technology and Open Source in Websites of Academic Libraries Worldwide. *The Journal of Academic Librarianship*, 42 (4), 319-328.

Witt, M. (2009). Institutional Repositories and Research Data Curation in a Distributed Environment Libraries Research Publications. Paper 104. http://docs.lib.psu.edu/lib_research/104