

***Grey Literature: Scholarly Communication in a Digital World***

Date: 23 August 2019

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

## **Repositories and Institutional Grey Literature**

### **Pingping Du**

Library, China University of Mining and Technology, Xuzhou Jiangsu, China.

E-mail address: ppu168@126.com

### **Yuke Li**

Shanghai Publishing and Printing College, Shanghai, China.

E-mail address: 119448932@qq.com

### **Bolou Huang**

Library, Stony Brook University, USA.



Copyright © 2019 by Pingping Du, Yuke Li and Bolou Huang. 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:**

*COAR has launched a research project called "Next Generation Repositories" since 2016, which focuses on promoting itself to be research-oriented, open and conducive to innovation, so that it can facilitate the collective management of academic groups. The traditional repositories as the management of research outputs (thesis, thesis, etc.), at first, are basically in the category of "white literature"; Secondly, the publication of management on the Internet (forums, academic blogs, etc.), basically belong to the "white literature" (The scope of White Literature). As the research data collection and management involved in the scientific research process is not fully integrated into the management domain of repositories, they are basically in the category of "grey literature". Research data set retention and long-term management is a part of the academic practice of the discipline, and the data management problem and incorporating it into the academic practice has become a hot topic in the academic circle. The Operation management of traditional repositories is beginning a new upgrade exploration - research data management (RDM). (1) the type and data structure of grey literature in the next generation repositories; (2) the relationship among "white literature", "like white literature" and "grey literature" in knowledge chain; (3) systematic analysis of the logical relationship of "study-data-result-structure-correlation-effect-multiplexing-openness". In order to reveal the multi-dimensional relationship between institutional knowledge resources and research data, the integration of RDM organization process data association law, the key elements of multi-source heterogeneous data, and the establishment of effective knowledge links are analyzed. (4) the*

*policy guarantee and implementation mechanism of the reuse and open utilization of institutional grey literature.*

**Keywords:** Repositories, Grey Literature, White Lite rapture.

---

## **1. Institutional knowledge database and its resources**

From the perspective of the university, Clifford A. Lynch defines the institutional knowledge database. He believes: "The institutional knowledge database is a set of services provided by the university for its employees to manage and disseminate the digital product created by its various departments and members". The director of the SPARC business unit Richard K. Johoson, considers it a collection of digital resources that capture and preserve intellectual products in single or multiple groups.

The main feature is: based on the network and the means of communication, the institutional knowledge database is built by an organization, for the purpose of collecting, sorting, preserving, retrieving, and providing, using various digital products created by members of the organization in the course of their work for its content. This is the traditional institutional knowledge database model, which mainly collects published papers, books, patents, reports, and so on.

The definition about institutional knowledge database of the National Science Library of the Chinese Academy of Sciences: "Institutional knowledge database is a tool for research institutions to implement knowledge management and to effectively manage their intellectual assets, and it is an important mechanism for institutional- knowledge- capacity building." In this definition, the scope of the institutional knowledge database extends from the organization's "knowledge products" to "knowledge assets". "Products" are products that are produced by themselves, while "assets" include not only their own products, but also the purchased knowledge products.

Therefore, the institutional knowledge database records, integrates, reorganizes, and displays the various types of knowledge stored in various systems, teams, and members of the organization. What is more, presenting organizational results, reusing results, and revitalizing institutional intellectual assets is to provide learning and research services support to the team and its members.

The asset range of the institutional knowledge database includes not only academic results, research process data, internal and external resources, and the use of resources.

## **2. Resource types of institutional knowledge resources & Definition of resource types**

Based on information sources, traditional literature is classified into black literature, grey/gray literature, and white literature. However, with the development of computer technology and information technology, the literature has developed into information, forming a diversified literature, information, and knowledge structure, and data has also become a resource type.

**Table 1: Resource types and features of institutional knowledge resource**

Types	Type extension	Connotation	content	Publicity	feature
White literature	White literature	Documents that are officially published and openly distributed among members of society, including books, newspapers, periodicals, etc.	Books, newspapers, periodicals, etc.	open	
	"white-like" literature	Other literature on social media communication		open	
Grey /Gray literature	Grey /Gray literature	Non-formal publications are generally referred to as grey/gray literature	Non-publicly published government documents, dissertations; non-publicly issued conference documents, scientific reports, technical files; corporate documents, enterprise product materials, trade documents (including product specifications, dynamic information materials issued by relevant institutions) that are not issued externally and Working documents; unpublished manuscripts and internal publications, exchange of materials, free reading materials, etc.	Selective disclosure	
	"grey/gray-like" literature	Various resources of literature value that have not been systematically processed in various corners and have not been able to appear in the form of similar databases.		Selective disclosure	
	data	Valuable knowledge resources formed during the research process	Scientific data, research data, scientific research data, etc.	Selective disclosure	
Black literature		Non-public publication, or a document with a narrow scope and confidential content.	Military intelligence information, technical confidential information, personal privacy materials, etc.	Not easy to disclose	

The resources of institutional knowledge resources are related ones generated by institutions and displayed in the form of traditional or digital resources, and the first generation of institutional knowledge base mainly collects white literature.

### 3. Resource composition of the institutional knowledge database

The construction and open sharing of characteristic resources of "knowledge capital level" are features of institutional knowledge database. Since 2016, COAR has launched a research project called "Next Generation Repositories", the core of which is to upgrade it to research-oriented, open and conducive to innovation, while facilitating the collective management of academic groups. Most of the institutional knowledge database built by university library in our country is a relatively independent "recombinant" static database, and its function is limited to the collection, collation, long-term preservation, retrieval and utilization of the open academic research results of the university. A few are related to the library management system, which has nothing to do with the information management system of the functional departments of the university, and the mining and preservation of some hidden resources is insufficient. It is not yet possible to realize the interaction, data reuse and value added between the whole life cycle data and users in the institutional knowledge base.

"Next Generation Repositories" study collection and preservation of data and academic results, the related organizations of multi-source and heterogeneous data, in particular, the organizational similarities, differences and fusion of research results and research data in the cross-domain of natural science research data and social sciences. The resource composition of "Next Generation Repositories" is more complete, covering the combination of the whole process knowledge chain of resource management, knowledge management, process management and other scientific research processes. Integrate "Next Generation Repositories" into the innovation system of knowledge resource recycling, reveal and discover the data of various innovation achievements and process results (including experts, knowledge, data, utilization, etc.) of the organization, in order to carry out organization, information and innovation management of knowledge resources for the organization, activity, achievement, development, effect and utilization.

**Table 2:Resource Composition of the Next Generation Repositories**

Stage,	Resource	Source
Researcher	Scholars/Scholar resources	Institution
Research process	Research data, scientific research data	Researcher
	Scientific data	Research institute
Research results	Papers, patents, books, etc.	
Knowledge achievement	Think-tank, etc.	
Result	Report, network	

### 4. The composition and management of the white literature and gray literature in the institutional knowledge database

Grey literature is called "grey" because these documents have certain concealment, but they do not have confidentiality and more publicity, which mainly circulated in each unit, have a large number of industry internal information and the potential function of sustainable development and utilization, and which do not have the concentration and systematicness of information. Grey literature mostly publishes the internal information of various units in various industries, the latest cutting-edge scientific and technological content, professional

construction, teaching planning, academic trends, research results and so on, which is more attractive than general academic journals. The most important grey literature in the new generation repositories should be scientific data, which is different from white literature and black literature. Therefore, we should first clear the scope of grey literature if we want to collect grey literature. The current grey literature mainly includes the following types.

According to the needs of special profession, about the “white-like” document type, the availability collection principle suitable for the construction of library characteristic database is formulated, that is, the white-like literature with certain academic, systematic and stable nature, which meets the needs of the database. It mainly includes the following types.

**Table 3: Grey resource composition of the next generation repositories**

Resources name	Form	Document definition
Scholars/Scholar resources	Scholar data/information	Gray/Grey
Research data, scientific research data, scientific data	Data during the research process Materials such as technology data, charts and original records, etc.	Gray/Grey
Internal publication	Internal magazines, promotional materials	Gray/Grey
Meeting Information	Meeting PPT, report draft, etc.	Gray/Grey
Think-tank report	Approval document, declaration	Gray/Grey
Teaching materials	Self-edited textbooks, handouts, subject construction materials, excellent course materials, excellent teaching plans, courseware, speeches, syllabus, teaching calendar, teaching method research, teaching evaluation, test questions and answers.	Gray/Grey
Research materials	Project report, completion report, and compilation of scientific research results, achievement appraisal, and patent application documents.	Gray/Grey
Undergraduate thesis	Thesis papers of undergraduate graduate are not collected or published. Most of these dissertations are scientific research summaries made by degree applicants under the guidance of tutors, which have a wide range of topics, rich in content, unique viewpoints, novel ideas, unique insights, and Has a large amount of chart data. The data has been academically valued by experts.	Gray/Grey
College students' scientific and technological innovation achievements	There are related materials and literature such as various competition-based innovation projects, invention patents, project papers, design works, speech programs, and various social practices, which are participated by students based on the knowledge they have learned.	Gray/Grey
Literature taken social media as a carrier	Various literature resources such as interactive blog articles, web forums, online original videos, online original music, encyclopedias, interactive questions and answers, etc. based on social network services.	White-like
Network electronic literature	Open access to periodicals, books, newspapers, degree papers are produced and disseminated and provided through the Internet; Those government official information, product introduction, education and teaching information, scientific research information, etc. are released to the public by government, enterprises, school research institutions and so on ; Papers, novels, essays written on the personal home page; Electronic pictures and animations taken and edited, and so on. From the public form, there are full public information, semi-public information, non-public information; from the storage format, there are document class format, audio class format, video class format, image class format, etc. From the point of view of software conditions, there are Email resources, BBS resources, Wiki resources and so on.	White-like

Multimedia video material	A series of static images are captured by electrical signals, recorded, processed, stored, transmitted and reproduced on the Internet. When the continuous image changes more than 24 (frame) pictures per second, according to the principle of visual temporary retention, the human eye cannot distinguish a single static picture ; It appears to be a smooth and continuous visual effect, but the video recording fragments can exist on the Internet in the form of streaming media and can be received and played by the computer with the development of network technology.	White-like
Courseware	It mainly refers to the open teaching resources and training resources created by institutions at all levels and colleges and universities around the world. These open curriculum resources can be downloaded free of charge through the network.	White-like
Picture (material)	Picture refers to the flat media composed of graphics, images and so on. The picture information published on the Internet, including social life pictures reflecting 100 forms of life, social documentary pictures that truly reflect people or things, natural landscape pictures and so on, can be selectively collected according to needs.	White-like
Industry news, conference information materials	The collection and processing of industry news and conference information is conducive to strengthening the exchange of academic information, forming the characteristics of their respective literature, and serving for the discipline construction.	White-like
Report, network		

## 5. Grey Resource Association Management and Reuse of Next Generation Repositories

At present, most of institutional knowledge database in colleges and universities are those of research results, and the resources of the construction are mainly academic achievements. The new generation of institutional knowledge database of colleges and universities is the expansion of the traditional ones, which is the co-construction of the scientific researchers, the research data produced in the scientific research process and the research results completed in the scientific research process. The problems to be solved in the construction are the collection and preservation of grey data, the related organization of relational data, the release management and so on. On the basis of the original institutional library, the main development and improvement mainly include: the expansion of the construction content of the new generation university institutional knowledge database; the collection of data in the whole process of scientific research; the specification of data preservation and the way of data preservation; the relationship and display of data in the process of scientific research: the sharing and open utilization of related data about/of scientific researchers, research data and research results.

The way of scientific data management in foreign colleges and universities has evolved three stages: from institutional knowledge base to data storage management center and then to data supervision and release center. At the beginning, the institutional knowledge base or institutional warehouse of colleges and universities was limited to the preservation of research papers, reports or grey literature and other results data., while the grey data set research data base has become a trend and development goal with the increasing demand for research data management and sharing in the public and scientific community. Generally speaking, colleges or universities and research institutions have more research data. As the process result produced by social investigation, scientific experiment, scientific exploration, data census and so on in the scientific process, it is the basis of scientific research and the conclusion fact certification. There are a lot of information in the research data, including the attribute, source, trend, the relationship between the data and so on, so it is necessary to

establish a scientific research system, sustainable research and secondary utilization of the data, and carry out the collection and archiving management of the research data. The attribute analysis and relationship correlation of scientific data can make readers understand the link relationship between data and data intuitively, and provide support for readers' scientific research topics or scientific research process. See the data relationship figure 1.

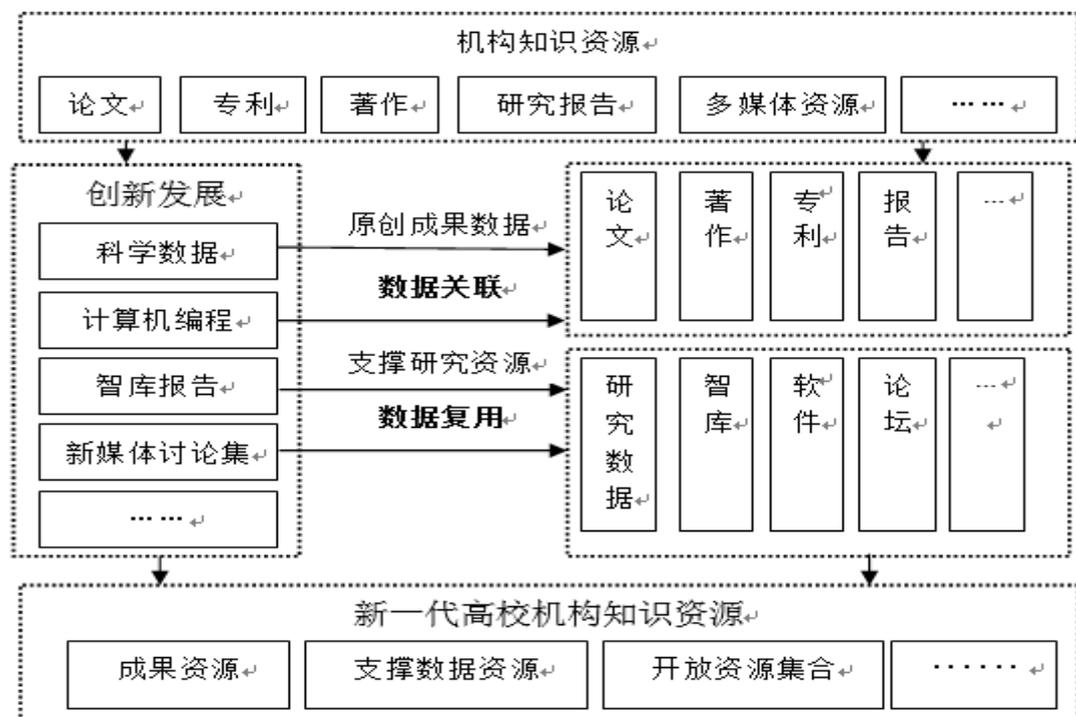


图1机构知识库和“新一代”机构知识库关联关系

figure1 the association relation between institutional knowledge database and "new generation" institutional knowledge database

(图中(机构知识资源 Institutional knowledge resources,论文 paper/ thesis 专利 patent 著作 book/literary work 研究报告 research report 多媒体资源 Multimedia (resources)创新发展 Innovation and development, 原创成果数据 Original results data, 论文 paper/ thesis, 著作 book/ literary work, 专利 patent, 报告 report, 科学数据 Scientific data, 计算机编程 Computer Programming, 数据关联 Data association, 智库报告 Think-tank report, 支撑研究资源 Support research resources, 新媒体讨论集 New media discussion set, 数据复用 Data reuse, 研究数据 research data, 智库 Think-tank, 软件 software, 论坛 forum, 新一代高校机构知识资源 A new generation of university institutional knowledge resources, 成果资源 Outcome resources, 支撑数据资源 Supporting data resources, 开放资源集合 Open resource collection)

The research on the association organization of data resources enables the digital resources of the institutional knowledge database to be effectively correlated and highly aggregated, so that they can fully meet the needs of users who are able to obtain the required resources quickly, accurately and conveniently. The associated data has the characteristics of high association, full sharing, easy expansion, convenient reuse and so on. As an important modern strategic resources, data resources will become more and more prominent. In this century, they may surpass oil, coal and minerals, and become one of the most important human resources.

## References

[1]Karasti H, Baker K S, Halkola E. Enriching the Notion of Data Curation in E-Science: Data Managing and Information Infrastructuring in the Long Term Ecological Research (LTER)

- Network[J]. *Computer Supported Cooperative Work*, 2006, 15(4):321-358.
- [2] Wu Jianzhong. From Open Access to “Beyond Open Access”—Enlightment from “Next Generation Repositories”[J]. *Journal of the National Library of China*, 2018, 27(06):3-8.
- [3] Managing Research Data.[EB/OL]. [2018-10-25]. <http://www.jisc.ac.uk/whatwedo/programmes/mrd.aspx>.
- [4] Wilson J A J, Martinez-Uribe L, Fraser M A, et al. An institutional approach to developing research data management infrastructure[J]. *International Journal of Digital Curation*, 2011, 6(2): 274-287.
- [5] Rice R, Haywood J. Research data management initiatives at University of Edinburgh[J]. *International Journal of Digital Curation*, 2011, 6(2): 232-244.
- [6] Latham, Bethany. Research Data Management: Defining Roles, Prioritizing Services, and Enumerating Challenges[J]. *The Journal of Academic Librarianship*, 2017, 43(3):263-265..
- [7] Zhang Jilong, YinShenqin, ZhangYong, et al. Social Scientific Data Sharing and Serving—An Example of Fudan University Social Scientific Data Platform[J]. *Journal of Academic Librarians*, 2015, 33(01):74-79.
- [8] GU Liping. Data Governance: Opportunity for the Library[J]. *Journal of Library Science in China*, 2016, 42(05):40-56.
- [9] Kuk G, Davies T. The Roles of Agency and Artifacts in Assembling Open Data Complementarities[A]. *Proceedings of Thirty Second International Conference on Information Systems[C]*, Shanghai, China, 2011.
- [10] Janssen M, Estevez E, Janowski T. Interoperability in Big, Open, and Linked Data--Organizational Maturity, Capabilities, and Data Portfolios[J]. *Computer*, 2014, 47(10):44-49.
- [11] SU Xinning. Opportunities and Challenges Faced by Digital Libraries in the Era of Big Data[J]. *Journal of Library Science in China*, 2015, 41(06):4-12.
- [12] Liu Wei, XiaCuijuan, ZhangChunjing. Big Data and Linked Data: The Emerging Data Technology for the Future of Librarianship[J]. *New Technology of Library and Information Service*, 2013(04):2-9.
- [13] Vandenbussche P Y, Atezing G A, Poveda-Villalón M, et al. Linked Open Vocabularies (LOV): a gateway to reusable semantic vocabularies on the Web[J]. *Semantic Web*, 2017, 8(3): 437-452.
- [14] Alonen M, Kauppinen T, Hyvönen E. Vocab.at—Automatic Linked Data Documentation And Vocabulary Usage Analysis [EB/OL]. [2018-10-29]. <https://seco.cs.aalto.fi/publications/submitted/alonen-et-al-vocab.pdf>.
- [15] Genetics N. European Open Science Cloud[J]. *Nature Genetics*, 2016, 48(8):821.
- [16] Tenopir C, Birch B, Allard S. Academic libraries and research data services[J]. *Library*, 2012, 46(3), 61–75.
- [17] Wang Yuefen, XieQingnan, SongXiaokang. Review and Prospect of Overseas Research on Data Science[J]. *Library and Information Service*, 2016, 60(14):5-14.
- [18] PLOS' New Data Policy: Public Access to Data.[EB/OL]. [2018-10-29]. <https://blogs.plos.org/everyone/2014/02/24/plos-new-data-policy-public-access-data-2/>.
- [19] Pasquetto I V, Randles B M, Borgman C L. On the Reuse of Scientific Data[J]. *Data Science Journal*, 2017, 16(8):1-9.
- [20] Huang Ruhua, WangBin, ZhouZhifeng. A Study on Countermeasures for Promoting Scientific Data Sharing in China[J]. *Library*, 2014(03):7-13.