

# Case Studies of Research Data Repositories in Humanities and Social Sciences

Wenying Liu  
Xiamen University Library  
Xiamen, China  
liuwy@xmu.edu.cn

Zheng Li\*  
Xiamen University Library  
Xiamen, China  
lz82@xujc.com

Jing Lin  
Xiamen University Library  
Xiamen, China  
linjingxmulib@xmu.edu.cn

**Abstract**—In order to explore research data management services in the field of humanities and social sciences by analyzing research data repositories we select three typical study examples: ICPSR Data Archive, UK Data Services and European Union Open Data Portal. This paper compare these three data repositories based on the registered information in re3data.org, including data content types, construction institutions types and openness of data repositories, etc. The United States, United Kingdom and European Union lead the construction and sharing of research data repositories in the field of humanities and social sciences. They have high protection awareness of research data, and have formulated corresponding policies at the national level. They attach importance to cooperation between organizations and are supported by funding agencies. These mature experiences can provide us with a good reference.

**Keywords**—*Research Data; Data Repositories; Research Data Management; Humanities and Social Sciences*

## I. INTRODUCTION

Research data is an important foundation for scientific work. Long-term archiving and accessibility of research data contribute to the traceability and quality of scientific work, enabling researchers to carry out other tasks.

This paper study base on re3data.org, which is a global registry of research data repositories that covers different academic disciplines. It presents repositories for the permanent storage and access of data sets to researchers, funding bodies, publishers and scholarly institutions [1]. Until March 20, 2018, there were 2037 data repositories registered at re3data.org, of which 601 were in the field of humanities and social sciences.

From these we select three typical examples for study: (1)ICPSR Data Archive -the largest social science data management project in the United States; (2)UK Data Services - the UK's national scientific data storage center and the UK's largest digital data collection institution for the humanities and social sciences; (3)European Union Open Data Portal - currently there are approximately 56 EU institutions, agencies or departments providing high-value public data in a wide range of EU policy areas (e.g. economy, employment, science, environment and education).

## II. EXAMPLES

### A. *Inter-university Consortium for Political and Social Research (ICPSR data archive)*

Founded in 1962 [2], ICPSR is an international alliance of more than 750 academic institutions and research institutes. It belongs to the University of Michigan, and the purpose is to provide scientific data acquisition, data management and data analysis methods in the field of social science research. Currently, ICPSR has more than 250,000 data archives of social and behavioral scientific research archives. It has collected 21 specialized data sets in the fields of education, aging, criminal justice, drug abuse, and terrorism.

1) ICPSR is responsible for managing and saving data to ensure that data resources can be used for a long period of time. Although many of the datasets in the ICPSR are freely available to the public, most datasets require users to become members of ICPSR member agencies or to pay for downloads.

2) Since 1963, ICPSR has provided training in quantitative methods to promote effective data use.

3) The ICPSR Summer Program in Quantitative Methods of Social Research provides a comprehensive curriculum for research design, statistics, data analysis and methodology.

4) In addition, ICPSR also provides user support to assist researchers in identifying relevant data for analysis and conducting research projects.

### B. *UK Data Service*

UK Data Service was established in October 2012 and is funded by the Economic and Social Research Council (ESRC) in the UK to meet the needs of researchers from all walks of life. Student and teacher needs [3]. The UK Data Service currently has more than 6,000 datasets, including: British census data, government-funded surveys (such as the UK Labour Survey and Health Survey), vertical studies (such as the UK Family Group Survey and the Millennium Study), and international Surveys (such as European Social Surveys and Euro Scales), international macro data (such as World Bank

and International Monetary Fund statistics), business microdata, and qualitative and hybrid methodological data.

1) Access data directories, documents and guides are free. Registration is required to download data, and data usage must comply with the licensing requirements set forth by the data owner.

2) Follow active protection policies to ensure the authenticity, reliability, and logical integrity of all digital resources while providing a permanently available version for research, teaching, or learning.

3) Encourage data sharing and reuse as an intrinsic value for expanding raw data for replication of research results and further analysis and teaching use.

*C. European Union Open Data Portal (EU ODP)*

According to the European Commission resolution (2011/833/EU) on the re-use of European Commission documents, all EU institutions are invited to provide their data as publicly as possible. In 2012, the European Union Open Data Portal was established. To date, approximately 56 EU institutions, agencies or departments (such as Eurostat, the European Environment Agency, and the Joint Research Centre and other European Commission Directors and EU institutions) have provided a total of more than 7.800 available data sets.

1) Allow anyone to easily search, browse, link, download, and reuse data through catalogs for commercial or non-commercial purposes. Users can access data stored on the websites of EU institutions, agencies and other agencies through the catalogue.

2) Semantic technology provides new functionality. The metadata catalog can be accessed through the interactive search engine (tab -Data) and SPARQL query (tab -Linked data). There are also visual application demonstrations from various institutions, institutions and other institutions in the European Union.

3) Users can suggest data they wish to link to, provide feedback on data quality, and share information with other users on how they use the data.

4) The portal interface is provided in 24 EU official languages, but most of the metadata currently only provide a limited number of languages (English, French and German), and some metadata (e.g. data provider names, geographic coverage) have 24 Language.

**III. CONTRAST**

We compare the three data repositories based on the registered information in re3data.org, including data content types, construction institutions types and openness of data repositories.

**TABLE I. DATA CONTENT TYPES**

	<i>Content Types</i>
ICPSR data archive	<ul style="list-style-type: none"> <li>① Standard office documents</li> <li>② Structured graphics</li> <li>③ Scientific and statistical data formats</li> <li>④ Plain text</li> <li>⑤ Structured text</li> <li>⑥ Archived data</li> </ul>
UK Data Service	<ul style="list-style-type: none"> <li>① Standard office documents</li> <li>② Structured graphics</li> <li>③ Scientific and statistical data formats</li> <li>④ Plain text</li> <li>⑤ Raw data</li> <li>⑥ Images</li> <li>⑦ Audiovisual data</li> </ul>
EU ODP	<ul style="list-style-type: none"> <li>① Standard office documents</li> <li>② Plain text</li> <li>③ Software applications</li> <li>④ Raw data</li> <li>⑤ Images</li> <li>⑥ Structured text</li> <li>⑦ Scientific and statistical data formats</li> </ul>

We can see from TABLE I that, standard office documents, scientific and statistical data formats , plain text are all include in three data repositories, other content types involve

structured graphics and text, archived data, raw data , audiovisual data, etc.

TABLE II. CONSTRUCTION INSTITUTIONS TYPES

	<i>Institution name</i>	<i>Institution responsibility type</i>	<i>Institution type</i>
ICPSR data archive	① List of Member institutions of ICPSR,	general funding	non-profit
	② University of Michigan, Institute for Social Research, Inter-University Consortium for Political and Social Research	general technical	
UK Data Service	① University of Manchester, Cathie Marsh Institute for Social Research,	general funding	non-profit
	② University of Essex, UK Data Archive,	general technical funding	
	③ Economic and Social Research Council (ESRC)	general funding	
EU ODP	European Union	general technical funding	non-profit

TABLE II list the construction agencies of the three data repositories. We can see that all the institutions are non-profit. From the aspect of responsibility type, they are all constituted

by general, technical and funding institutions. All construction agencies are multi-attribute, for example, European Union is a general and technical institution, and also a funding institution.

TABLE III. OPENNESS OF DATA REPOSITORIES

	<i>Database access</i>	<i>Data access</i>	<i>Data upload</i>
ICPSR data archive	open	open; restricted; closed	restricted
UK Data Service	open	open; restricted; closed	restricted
EU ODP	open	open	restricted

The openness of data repositories can be reflected in three aspects: database access, data access and data upload, and access levels are divided into: open, closed, and restricted (TABLE III). Fully open means that there are no access barriers; restricted use means that external users can overcome access barriers; closed means that external users cannot overcome entry barriers.

From the perspective of database access permissions, the three databases are all completely open to public; judging from the data access rights, data is completely open in EU ODP, and only part of data in ICPSR data archive and UK Data Service are open, while rest are restricted or closed; from the perspective of data upload permission, all the data are restricted to upload (have to become member or pay for it).

IV. CONCLUSION

The United States, United Kingdom and European Union lead the construction and sharing of research data repositories in the field of humanities and social sciences. Through the analysis of typical cases, we could find the reasons for the prosperity of research data repositories construction in developed countries.

1) High protection awareness of research data is a prerequisite for data management. Today, opening up has gradually begun to become a global consensus. From "open access" to "open data" to "open science", this open movement characterized by knowledge sharing continues to develop in depth.

2) Data policy is the basis for data sharing. In 2011, the European Commission promulgated the report “Open data: An engine for innovation, growth and transparent governance”<sup>5</sup>, which requires the EU and its member states to establish legal mechanisms and adopt fiscal measures to promote cooperation among countries in the field of open data. The United States, the United Kingdom, etc. have also issued policies to promote data opening at the national level.

3) Funding agencies play a leading role in policy formulation. For example, Economic and Social Research Council (ESRC), which found UK Data Service, requires applicants to submit a data management plan as part of their Je-S application; all data created or reused during the ESRC authorization must be provided within three months of the end of the authorization for reuse or archiving, etc<sup>6</sup>.

4) A good institutional cooperation atmosphere is an important factor in promoting open data. Data management is a huge project that is difficult to accomplish with only one effort. The three data repositories covered in this article are jointly established by various types of agencies, including general, technical, and funding institutions.

#### REFERENCES

- [1] re3data.org. [EB/OL]. [2018-03-20] .<https://www.re3data.org/about>
- [2] ICPSR. [EB/OL]. [2018-03-20]. <http://www.icpsr.umich.edu/icpsrweb/content/about/>
- [3] Ukdataservice [EB/OL]. [2018-03-20] .<https://www.ukdataservice.ac.uk/about-us>
- [4] European Union Open Data Portal. [EB/OL]. [2017-10-23]. <https://data.europa.eu/euodp/en/data>
- [5] Communication Commission. Open data:An engine for innovation, growth and transparent governance. [EB/OL] . [2018-03-20]. [http://www.europarl.europa.eu/RegData/docs\\_autres\\_institutions/commission\\_europeenne/com/2011/0882/COM\\_COM\(2011\)0882\\_EN.pdf](http://www.europarl.europa.eu/RegData/docs_autres_institutions/commission_europeenne/com/2011/0882/COM_COM(2011)0882_EN.pdf)
- [6] Research Data Policy. [EB/OL] [2018-03-20]. <http://www.esrc.ac.uk/funding/guidance-for-grant-holders/research-data-policy/>.