ICSSR Data Service: Policy Guidelines

1. Introduction

Researchers in every discipline, from the arts and humanities to engineering science and technology produce a large amount of research data. Engineering and Physical Sciences Research Council (EPSRC, 2015) defines research data as "recorded factual material commonly retained by and accepted in the scientific community as necessary to validate research findings; although the majority of such data is created in digital format, all research data is included irrespective of the format in which it is created". This process covers all stages of research, ranging from research data generation, which may be any kind of experiments in the sciences, a study in the social sciences and so on. Research data can be raw data produced in the laboratory during a scientific experiment or through a survey, or it can be processed data that has been cleaned, refined, arranged, and combined in a manner that is useful in research. Research data also include data that has already been published in the journals or in other scientific communications. Research data includes analogue sources as well as discrete digital objects (text, files, image, audio, and video), complex digital objects (discrete digital object made by combining a number of other digital objects) and databases (Whyte and Tedd, 2011).

Data itself has its own value as an output of research, enhancing the profile of researchers and institutions. Researchers require physical, computing and network infrastructure that ensure the maximum accessibility, stability and reliability to facilitate working with (reuse) and sharing of their research data. There has been a persistent demand across the globe for open access to research data collected or developed through public funding. Open data provides for its free availability, preservation, long-time sharing and long-term use.

According to National Data Sharing and Accessibility Policy (NDSAP, 2014), “A dataset is said to be open if anyone is free to use, reuse, and redistribute it – Open Data shall be machine readable and it should also be easily accessible.”

In the background given above, several leading organizations worldwide are coming together to work towards the goal of data sharing, preservation, accessibility and reuse. In the year 2003, “Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities” pointed out the importance of open access to sciences and humanities research data as well as heritage data and the relevance of research data as an integral part of scholarly knowledge. The Organization for Economic Co-operation and Development (OECD) published “Principles and Guidelines for Access to Research Data from Public Funding” in 2007 with an aim to promote data access and sharing among researchers. The Royal Society in UK has also appealed to the scientists to make their data accessible and usable through an appropriate data repository in their report entitled “Science as an Open Enterprise” in the year 2012. In 2013, the U.S. Government’s Office of Science and Technology Policy (OSTP) issued directives stating that digitally formatted scientific data resulting from unclassified research supported wholly or in part by Federal funding should be stored and publicly accessible to search, retrieve, and analyze.

2. Data Sharing and its Importance

Data sharing is the practice of making data used for scholarly research available to other investigators / researchers for reuse. Several funding agencies, institutions, and publications have well-defined policies for data sharing because transparency and openness are increasingly considered to be important part of the scientific process. The value of any research data lies in their use and reuse. In addition, publicly funded research data are produced in the public interest and, therefore, should be shared widely. The

concept of “open data” and data sharing is encouraged as a policy because of the reasons mentioned below:

- Data sharing enhances the reuse of data for further scientific and educational purposes;
- It promotes the impact and visibility of the data, and its associated research projects;
- Data sharing helps to promote the profile and recognition of the researchers, institutions and research funders to a great extent;
- Sharing of data maximizes the possibility of data transparency and accountability for a completed research and its output;
- The practice of sharing of data allows others to understand, evaluate and advance a piece of research while relying on an individual / group of individuals or organization’s specific work, and thereby providing them an opportunity to collaborate in further research on the same topic;
- Data sharing encourages scientific inquiry and debate; and
- It also helps to check the duplication of an existing research work thereby helping the policy makers and research funders.

2.1. Data Sharing Methods

Major routes of sharing scientific and research data are as follows:

- **Data Repositories**: Repository refers to a central place where data is stored and maintained, often for safety and preservation. It can be a place where multiple databases, datasets or files are located for preservation and distribution over a network that is directly accessible to the user. It may include requisite infrastructures, often referred to as data archives or data centres, required for obtaining

and depositing data to facilitate further sharing, analysis and reuse. Data repositories are continuing to evolve in almost all disciplines as an active component of virtual research ecosystem.

- ☞ **Institutional Repositories**: Usually based in an academic institution, institutional repositories store and share data created by a given institution's faculty members, staffs or students.
- ☞ **Self-Preservation**: Self-preservation is the only option available to an individual researcher to share data by publishing it or through his / her website or on discussion forums.

3. Research Data Management Plan (RDM)

A well-defined Research Data Management (RDM) is a basic necessity for efficient access, share and reuse of data. The RDM deals with organisation of data, from its entry into the research cycle to its dissemination and archiving of valuable results. It aims to ensure reliable verification of results and permits new and innovative research built on existing information (Whyte and Tedds, 2011). Research Data Management ensures effective management of data and its long-term availability. It deals with the development and implementation of practices, procedures, and policies to protect, validate, and describe data at various stages of its life cycle. RDM promises rewards for data sharing throughout and beyond the life of a research project. It ensures that data are discoverable and accessible to enable long-term preservation and reuse starting from its creation (research-active phase). Most developing countries do not have proper policy guidelines for an RDM either at national level or institutional level.

It is necessary to consider the Research Data Life Cycle to formulate effective policies and guidelines for sharing research data and successful implementation of RDM plan.

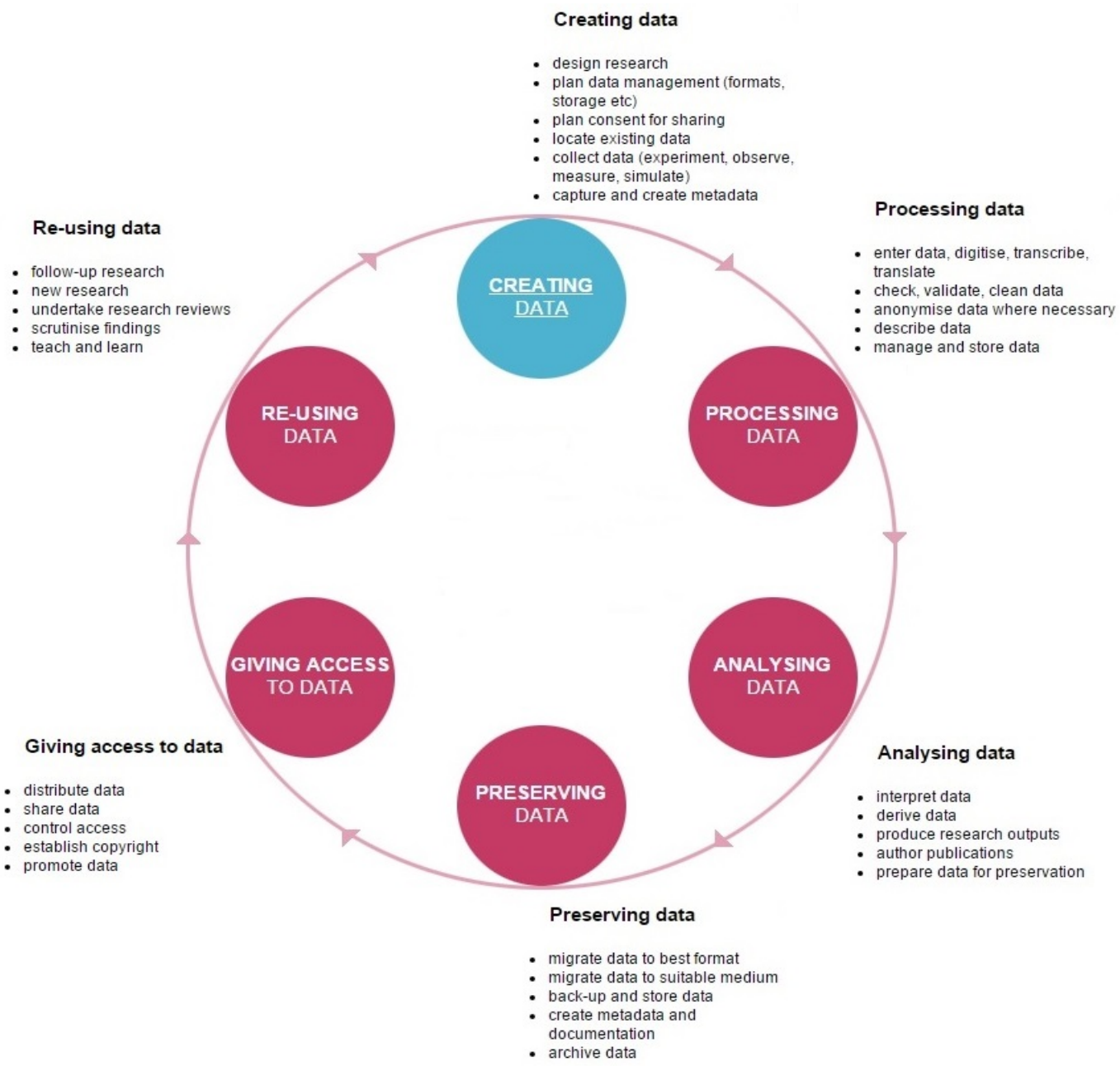

Figure-1: Representation of Research Data Life Cycle produced by UK Data Archive

4. Data Sharing in Indian Scenario

The Government of India has committed itself for data sharing through a policy document entitled "The National Data Sharing and Accessibility Policy (NDSAP)" published in the Gazette of India in March 2012

with an aim to share non-sensitive data available either in digital or analogue format that is generated using public funds by various ministries / departments / subordinate offices / organizations / agencies of Government of India as well as states. The NDSAP policy is designed to promote data sharing and enable access to Government of India owned data for national planning, development and awareness. Based on the NDSAP Policy, the flagship initiative "data.gov.in" of the Government of India was launched on the Open Data Government (OGD) platform in the year 2012 to act as single-point access to all resources (datasets / apps) in open format published by various ministries / departments / organizations of Government of India.

In accordance to Petroleum and Natural Gas Amendment Rules 2006, the Directorate General of Hydrocarbons (DGH), nodal agency under Ministry of Petroleum and Natural Gas, Government of India has started to develop a National Data Repository (NDR) to preserve different kinds of data i.e. oil-field data, cultural data, geological data, petro-physical data, seismic data, well data, production data, reservoir data, and various unstructured data (reports, documents, etc.). The NDR has started collecting relevant data to store and disseminate for its sharing and reuse.

The Indian Council of Social Science Research (ICSSR) is an apex body of Government of India that sponsors, promotes, monitors social science research and provides financial help to carry out research in the country. With an aim to make all the social science and statistical data/datasets freely available in open access to entire social science research community not only in India but all over the world and to foster substantial social science research in the country.

The "ICSSR Data Service" is culmination of signing of Memorandum of Understanding (MoU) between Indian Council of Social Science Research (ICSSR) and Ministry of Statistics and Programme

Implementation (MoSPI). The MoU provides for setting-up of "ICSSR Data Service: Social Science Data Repository" and host NSS and ASI datasets generated by MoSPI. Under the initiative, social science research institutes, NGOs, individuals and others dealing with social science research are also being approached to deposit / provide their research datasets for hosting into the repository of ICSSR Data Service. The ICSSR Data Service includes social science and statistical datasets of various national-level surveys on industries, employment and un-employment, household consumer expenditure, enterprise, land holdings survey, census data, etc. into its repository.

ICSSR Data Service aims to facilitate data sharing, preservation, accessibility and reuse of social science research data collected from entire social science community in India & abroad. The Information and Library Network (INFLIBNET) Centre, Gandhinagar has been assigned the task of setting-up the data repository.

5. Policy Guidelines for Research Data available in Social Science Data Repository (ICSSR Data Service)

The policy guidelines is prepared as a source for individual researchers, institutions and others who wish to or plan to deposit and share their research data through the ICSSR Data Service: Social Science Data Repository. The policy guidelines cover several aspects of research data from its creation to its sharing and re-use based on effective RDM plans which are in practice around the world. These guidelines are outlined below:

5.1. Content Coverage

5.1.1. Scope: Subjects and Languages

The scope of ICSSR Data Service is to accommodate data collected from social science and other related subject domains.

Currently, English is the basic required language to collect for and deposit data in the data repository. Data in other Indian languages may be considered in due course of time.

5.1.2. Kinds of Research Data

The data repository includes social and statistical data of various national-level surveys on industries, education, employment and un-employment, household consumer expenditure, enterprise, land holdings survey, census data, etc. Also, it collects quantitative and qualitative scientific research data generated from various social science institutions, research centres and others.

Supplementary information or materials associated with the research are generally included in the data repository. These accompanying materials may be interviewer guides, questionnaires, data collection methods, codebooks, database dictionaries, project summaries / descriptions, and bibliographies of publications pertaining to the data, etc.

5.1.3. Status of the Research Data

The process of data generation in a scientific research passes through several phases in its life-cycle, i.e. from raw data generated directly from scientific instruments or data collected through surveys to selection, validation, standardization, and other filtering and curation process.

The ICSSR Data Service determines the status of data to qualify it for inclusion in the data repository. It considers mainly:

- ☞ the raw or preliminary data;
- ☞ data that are ready to use;
- ☞ data sets that are ready for full release;
- ☞ unit-level summary data; and
- ☞ tabulated, analyzed and derived data, etc.

The ICSSR Data Repository mainly holds the unit-level statistical data of National Sample Survey (NSS) conducted by National Sample Survey Office (NSSO), and Annual Survey of Industries (ASI) conducted by Central Statistics Office (CSO) - Industrial Statistics (IS) wing along with their supporting documents, i.e. questionnaires, data collection methods, codebooks, and project summaries / descriptions, etc.

5.1.4. Version Control

It is necessary to keep various versions of a dataset along with the original deposited dataset in the repository as dataset can be updated, edited, revised and deleted. A version control policy clearly demonstrates and ensures the originality, authenticity and legal aspects associated with a datasets. ICSSR Data Service considers the following facts for the datasets available in the repository:

- ☞ The repository uses numbers for all versions of a dataset as per its updation, edition, revision, and also with its status, i.e. draft, interim, final, etc. to clearly indicate to their users.
- ☞ The same details may be recorded in DDI metadata record with the data file.
- ☞ The repository always keeps the original deposited datasets along with its supplementary information.
- ☞ A dataset is being kept in different formats along with its original within the repository.

- ☞ The repository uses Persistent Identifiers' to cross-link and cross-reference different versions of a dataset with the original.

5.1.5. Data File Formats

The data repository supports varieties of preferred data formats that are available for deposit into the repository which are machine readable. The unit-level NSS and ASI statistical data are available mostly in Excel, CSV, SPSS, and STATA formats. While, the supporting documents, i.e. questionnaires, data collection methods, codebooks, and project summaries / descriptions, etc. are in plain text, Word, PDF, and HTML formats. However, users have the liberty to export / download, convert and use the datasets using different statistical software packages according to their requirements.

5.1.6. Volume and Size Limitations

Currently, there are no restrictions on file sizes, or number of files that can be deposited into the repository. However, the data repository reserves the right to refuse or accept datasets or accompanied materials. The repository may discard datasets and materials that are too voluminous and may make it difficult or impossible to process given the resources, facilities or capacities of the repository's data operation. As and when necessary, the datasets / materials may be deposited and stored in compressed format in the repository.

5.2. Metadata

Metadata is the information that describes the data source and the time, place, and conditions under which the data were created. Metadata informs the user of who, when, what, where, why, and how data were generated. It is the data about data. Metadata allows the data to be traced to a known origin and

know quality. As such, all datasets submitted to the ICSSR Data Service should accompany with necessary metadata from the data owners (Data Producer / Depositor). Besides metadata, the datasets should also be accompanied with supplementary materials such as interviewers guide, questionnaires, data collection methods, codebooks, database dictionaries, objectives, project summaries / descriptions, etc.

5.2.1. Access to Metadata

Access to metadata records of the datasets available in the data repository is free of charge and openly accessible to all users. However, access to some of the metadata may be controlled or restricted.

5.2.2. Reuse of Metadata

Use and reuse of metadata is governed by policies and procedures of the ICSSR Data Service. Metadata can be accessed and explored online freely. The data and metadata available through the repository should be used only for the purposes of teaching and research, secondary research and analysis not for any commercial purposes. Proper attribution and citation should be given after use / re-use of metadata. Metadata records can be harvested for data sharing using OAI-PMH protocols, after seeking proper permission from ICSSR Data Service.

5.2.3. Metadata Types and Sources

The data deposited by researchers should be accompanied by detailed metadata documentation that describes the data, the datasets in it and the processes used to create it, etc. The data deposited into the ICSSR Data Service are described at the follows four levels:

- **Document Description**: Covers information on documents, i.e. its title, the data collector, data depositor and data producer, etc.
- **Study Description**: Information on data, scope of the study (topics, geography and time), methodology of data collection, sampling and processing, data access information, and information on accompanying materials, keywords, data citations, etc.
- **Datasets Description**: Information on data format, file type, file structure, missing data, weighting variables, keywords, data units and labels, etc.; and
- **External Resources Description**: Information about the external resources or supporting documents used for the research study and deposited with the data into the repository.

5.2.4. Metadata Schemas

The ICSSR Data Service employs DDI-XML-based descriptive metadata schema for items deposited into the repository. The Data Documentation Initiative (DDI) is a metadata standard developed by and for the social science data repositories and is a natural choice as a common metadata standard for the social science data infrastructures around the world.

Also, the Data Service assigns keywords to the data / datasets from Humanities and Social Science Electronic Thesaurus (HASSET) which is the leading British English thesaurus for social sciences developed by UK Data Service.

Example of a MoSPI Dataset's DDI Catalogue Record available in ICSSR Data Service with some mandatory and optional metadata elements are as follows:

Name of the Field	Element code	Value
Identification No	<IDNo>	DDI-IND-CSO-ASI-1983-84
Title	<titl>	Annual Survey of Industries (Detail) 1983-84
Alternate title	<altTitl>	ASI 1983-84
Series Name	<serName>	Industrial Statistics (Organised Manufacturing & Labour Sector) Survey
Depositor	<depositr>	--
Authoring Entity	<AuthEnty>	Central Statistics Office (Industrial Statistics Wing)
Data Collector	<dataCollector>	NSSO(Field Operation Division)
Funding Agency	<fundAg>	MOSPI, Government of India
Grant No	<grantNo>	----
Other ID	<othId>	Standing Committee on Industrial Statistics
Producer	<producer>	Field Operation Division, NSSO
Topic Classificatio	<topcClas>	Public Sector
Abstract	<abstract>	Abstract of data
Time Period	<timePrd>	1988-1992
Collection Data	<collDate>	01-07-1984
Nation	<nation>	India
Geographical	<geogCover>	Coverage of the Annual Survey of Industries extends to the

Coverage		entire Factory Sector, comprising industrial units (called factories) registered under section 2(m)(i) and 2(m)(ii) of the Factories Act.1948, wherein a "Factory", which is the primary statistical unit of enumeration for the ASI is defined as
Geographical Unit	<geogUnit>	Village, Taluk, District, State, National
Unit of Analysis	<anlyUnit>	Individual, Household, Factories etc
Type of Data	<dataKind>	Census and Sample survey data
Universe	<universe>	The survey cover factories registered under the Factory Act 1948.
Time Methodology	<timeMeth>	Cross-sectional (one-time) study
Sampling Procedures	<sampProc>	There was no deviation from sample design in ASI 1983-84.
Collection Mode	<collMode>	Statutory return submitted by factories as well as Face to face
Weightage to data	<weight>	(i) If Scheme Code = 1 then Multiplier = 1, If Scheme Code = 2 then Multiplier = 2 (ii) During Processing/Tabulating apply the multiplier to each characteristics
Sources of Data Collection	<sources>	Sources of data collection
Keywords	<keyword>	Transport Charges, Raw Materials
Date of	<distDate>	----------

Distribution		
Version	<version>	Version1.00: Reorganised Anonymized dataset for publication
Copyright	<copyright>	ASI 1983-84, CSO(IS Wing), Kolkata
Bibliographic Citation	<biblCit>	Central Statistics Office (Industrial Statistics Wing), MOSPI, Government of India. Annual Survey of Industries: 1983-84 [computer file]. × Edition. Kolkata, India: ICSSR Data Service [distributor], October 2012. SN: ×, DOI/Handle: xxxxxxxxxx
Access Restriction	<restrctn>	Open or Restricted
	<accsPlac>	---
Contact	<contact>	Contact information
Related study	<relStdy>	Related study
Publications related to present survey	<relPubl>	Publications related to present survey
Questionnaire	<qstn>	Questionnaire
Other reference materials	<othRefs>	Other reference materials

5.3. Submission of Data (Ingest)

5.3.1. Eligible Depositors

- **MoSPI**: Ministry of Statistical Planning and Implementation, Government of India;
- **ICSSR**: Indian Council of Social Science Research (ICSSR) and its 27 constituent research centres located across the country;
- **Universities and Academic Institutions**: Public-funded Indian universities, academic and other scientific research institutions; and
- **Others**: Any other organization or individual who may wish to deposit their data into the repository by complying with the data repository policies. These contributors may be individuals / independent researchers, government organizations, private firms, NGOs or any other institutions working in social science discipline and related domains.

All the depositors of data may only deposit their own work. Basically, there are two ways for data-owners to deposit their research data into the data repository, i.e. online or offline. Data owners may upload datasets online with all its supporting documents in appropriate format(s) into the repository along with the filled and signed copies of prescribed Data Deposit Form and Licence Agreement Form, or else they may sent their data as an attachment to the e-mail to the designated e-mail at ICSSR Data Service. Data owners may have the option to send their data offline too using postal or courier services to the "Data Collection Team, ICSSR Data Service".

5.3.2. Moderation by Repository

Acceptance of data may not compulsorily be guaranteed after submission as the data have to be internally evaluated by the data experts from ICSSR Data Service. Datasets proposed for deposit are reviewed for eligibility of the depositor, relevance to the scope of the collections, valid data formats, etc. Further, the submitted data would have to go through an internal data quality check to ascertain that

the quality standards are maintained, i.e. accuracy, its consistency, documentation, metadata, free from any sort of legal issues, privacy of individuals are maintained and data does not compromise with the national security. The submitted data would qualify for deposition into the repository only after thorough examination. The repository adds additional keywords and persistent identifiers' to each datasets deposited into it.

5.3.3. Data Quality Requirements

5.3.3.1. Responsibility

The ICSSR Data Service has a clear and concise Data Deposit Form and Licence Agreement Form written in simple language for the depositors of data which has to be submitted along with the datasets at the time of submission of data into the repository. In most cases, the depositors of data (Data Producers) are responsible for the quality of their research data. The ICSSR Data Service will be responsible for safe storage, availability, access and sharing of data available in the repository.

The ICSSR Data Service accepts no responsibility for mistakes, omissions, or any kind of legal infringements of data deposited into the repository. The data owners or data producers remain the rightful owner of the copyright of their data deposited in the repository.

In few cases, the Data Service may provide licenses to depositors to cover the limit of requirements for reuse of the data or vice-versa for the obligations on the part of end-users. The license includes the conditions under which access to the data may be given to the users, as specified by the depositor. Further, it states that the depositor has cleared all necessary permissions associated with the data to facilitate access on the part of the repository.

5.3.3.2. Quality Assessment

It is the sole discretion of the ICSSR Data Service to make decisions about whether to accept a dataset or not. The Data Service reserves the right to evaluate data quality in order to consider whether or not a dataset is useful for reuse through secondary analysis, and has a long-term value to it. As per the quality assessment check policy of the ICSSR Data Service, datasets submitted to the repository should be accompanied by a clear and complete documentation to enable end-users to understand the data contents and its reuse. Whether, the required contextual information (metadata) has been provided by the data producer or not at the time of submission of data it has to be created by the repository. The datasets along with the data files are duly checked and validated to ensure that variables and values are accurate according to the documentation supplied by the depositor and are sufficiently labelled for secondary use.

5.3.4. Confidentiality and Disclosure

Protection of identities of human subjects is of paramount importance in research practice and is an important ethical obligation of everyone involved in research process. Thus, it is too a sensitive issue on the part of ICSSR Data Service to ensure that the respondent's personal identification or information remain confidential when data are shared or made available through the data repository. In most cases, the ICSSR Data Service may require that the depositors of data (Data Producers) ensure that the data must meet requirements of confidentiality and non-disclosure for data collected from human subjects at anytime of research or at any steps of research data life-cycle. However, the data repository undergoes stringent confidentiality reviews to determine whether the data / dataset contain any direct or indirect identifiers that could be used / misused by third parties - on its own or in combination with other

publicly available information - to identify respondents. Some of the identities that are checked are as follows:

- **Direct identifiers** include but are not limited to names, addresses, telephone numbers, driver's license or Aadhar ID, etc.
- **Indirect identifiers** are items which, when used with other variables in the dataset, may provide enough detail to identify the respondent. It includes geographic detail like zip code, block number, etc.; membership nos. in clubs, groups, organizations, etc.; names of schools attended by the respondent or respondent's family; job titles, positions held in organizations, elected offices, etc.; and personal information such as income, events, etc.)

The ICSSR Data Service may alter sensitive data or may delete direct / indirect identifiers from the datasets to create anonymized data after consultation with the data investigator or producer of data or to create public use files that may limit the risk of personal disclosure. Only after the completion of those reviews, the data are made available from the repository to the users' community.

5.3.5. Embargo Period

In context of data repositories, the "embargo period" refers to a time delay, if any, of up to a maximum of 12 months following the submission of data until that is made available on the repository for the users' community. This period is set by the ICSSR Data Service for each datasets deposited into the repository based on the data depositor's claim in the Data Deposit Form or upon mutual agreement.

In few circumstances, it may be necessary for the ICSSR Data Service to limit access to a particular dataset for a specific time period not exceeding one year. These might include copyright restrictions,

sponsor's compliance, and research data management policies established by a particular research community for its research data. Further, the ICSSR Data Service may consider and decide whether to hold a given dataset and to what extent. In certain cases, the metadata may be made publicly accessible but the data is embargoed or restricted in some ways.

5.3.6. Rights and Ownership

The ICSSR Data Service needs a clear and concise statement from the producer of data as to who owns the data before it can be made available through the repository. The data owners are required to sign a license agreement with the ICSSR Data Service at the time of submission of data. As per the licence agreement, data owners and their institutions hold the copyright of the research data generated by them and deposited into the repository whereas non-exclusive rights is granted to the ICSSR Data Service to preserve and redistribute data through its data repository. Data will remain the property of the data owner / institutions / Government department / entity which collected them. However, the ICSSR Data Service ensures that the personal and private information about the data owner would not be disclosed in any manner. User must agree with the declaration of confidentiality of data and depositor.

If any copyrighted materials (i.e. scientific instruments or scales, etc.) were used by the data owners at the time of data collection and analysis, that should be clearly mentioned in the agreement so as to avoid any kind of legal infringements on part of the ICSSR Data Service in future. However, the ICSSR Data Service may help the data owners to obtain permission in such cases before disseminating such data through the repository. Moreover, the ICSSR Data service would also like to ensure that instruments used or scales along with data should be distributed under "fair use" and may not be re-distributed by users. Datasets available in the data repository are solely for secondary research as well as for research

and teaching purposes. The data or its derivatives should not be used for any kind of unfair and commercial benefits by any user.

5.4. Access and Reuse of Data

5.4.1. Access to Data Objects

The ICSSR Data Service promotes the philosophy of open access and adheres to open data policy. The ICSSR endorses the philosophy that "Digital research data should be easy to find, and access should be provided in an environment which maximises ease of use, provides credit for and protects the rights of those who have gathered or created data, and protects the rights of those who have legitimate interests in how data are made accessible and used (RIN, 2008)."

Datasets and their accompanying materials available in the ICSSR Data Service are generally publicly available. However, restrictions may be applicable on access and reuse of some specific datasets due to certain circumstances. The ICSSR Data Service facilitates access to each datasets deposited in the repository based on the request of the data owners at the time of signing license agreement. The following three types of access control can be exercised on the datasets available in the repository:

- **Open Data**: Access to open data generally means that data generated through public-funding should be freely available in open access in easy, timely, user-friendly and web-based methods without any process of registration / authorization to one and all without any restrictions.
- **Safeguarded Data**: Data can be access only to the registered users only after a prescribed registration process or authorization by the ICSSR Data Service once they agree to the terms and conditions of data usage policy displayed to them at the time of request.

- **Controlled Data**: Access to these kinds of data would be controlled through Secured Lab, stored in a secured server. Different versions of same datasets may be derived for its open and controlled access. Data declared as sensitive and highly confidential, by Government of India policies, will be accessible only through this mode.

5.4.1.1. Registration

It is necessary for all users to register so as to get access to the data available in the data repository except the open data. After registration, the users are asked provided a log-in ID & password to use the repository and download datasets. However, all users should read the terms and conditions of access and use of datasets carefully, then agree to the online Terms of Use statement available in the repository before downloading.

The process of registration to ICSSR Data Service is same for all users. Users have to go through the Registration Process to register themselves.

5.4.1.2. Search and Access

Datasets available in the repository are free to access and reuse for further study, research, teaching and learning. These datasets are well-organized according to their data characteristics. Users can search these datasets using search interface available on the ICSSR Data Service website to find out or locate their desired data/information. Users can also search the Data Catalogue which allows them to search by title, creator, keywords, description, subject, coverage, series, or identifier, etc. After finding their desired data, users may analyse the data online, export / download datasets for further self-analysis or

else may download the excerpts of their online analysis. Visualization tools, mapping and geo-coding applications are in-built in the data repository too.

5.4.2. Use and Reuse of Data Objects

5.4.2.1. Use and Reuse

The datasets deposited into the ICSSR Data Service are available in the public domain. As such, there are no restrictions or limitations on access and use / reuse of data for fair and personal use. Users are strictly prohibited to use the data for any kinds of commercial benefits. Misuse of data, data modification, repackaging, and redistribution of datasets in other formats are not permitted. User should not breach the intellectual property rights of the data owner while reusing the data.

5.4.2.2. Citation

It is a healthy academic and research practice to provide information regarding sources of data used in subsequent works. Citation is a significant feature of any research publication. The ICSSR Data Service expects that users of data from the repository should provide correct citation and acknowledgement for data used.

Example of a dataset from the repository is as follows:

Central Statistics Office (Industrial Statistics Wing), MOSPI, Government of India. Annual Survey of Industries: 1983-84 [computer file]. × Edition. Kolkata, India: ICSSR Data Service [distributor], October 2012. SN: ×, DOI/Handle: xxx xxxxx xxx

User may refer to: How to Cite Data? for further details.

5.4.3. Tracking Users and Usage Statistics

The most potential users of data are within the higher education and social sciences research communities though data may also be used by researchers from cross-sectional / sectors such as central & local government, private firms and researchers, etc. ICSSR Data Service provides usage of data for those datasets that require user registration to access. The data owners or depositors may write to the ICSSR Data Service if they wish to know the detailed usage of their data deposited in the repository.

5.5. Preservation of Data

5.5.1. Retention Period

In general, the ICSSR Data Service does not permit datasets to be withdrawn or removed once deposited into the data repository unless otherwise clearly specified by the data owners or depositors at the time of submission of data. The research data has to be retained for long term in order to be accessible for reuse and analysis in new research.

5.5.2. Functional Preservation

Data access, storage and preservation system for the ICSSR Data Services is based on National Data Archive (NADA) software developed by the International Household Survey Network IHSN) with dedicated production and storage server installed at INFLIBNET Centre, Gandhinagar, Gujarat. The Data Centre and the storage media used in a secured area with tightly controlled measures to avoid any kind of data loss in time of natural disaster or any other environmental vulnerability. Moreover, the ICSSR Data Service is mirrored at the ICSSR Head Office located at New Delhi. However, the ICSSR Data Service

will not be responsible for any loss of data or discontinuation of service in case of any events of force majeure.

5.5.3. File Preservation

The ICSSR Data Service store and preserve datasets in various machine readable file formats (*Please, see 5.1.5. Data File Formats*). The ICSSR Data Service follows established best practices for managing and preserving datasets for long term. The ICSSR Data Service cannot guarantee preservation of all datasets and metadata of those deposited into it for operating lifespan of data repository. Because, the readability of some of the currently deposited file formats may not be compatible in future due to software and technological obsolescence. However, repository may promise to maintain the usability of the specific file formats over time by data conversion, curation or creating those in other relevant formats for its interpretation. Datasets may be stored in compressed formats using proprietary compression software with all accompanying materials. The original submitted data files may be copied in different formats and stored in different spaces in the storage media. The ICSSR Data Service will regularly back up all the datasets in accordance with established best practices.

5.5.4. Integrity, Fixity and Authenticity

The ICSSR Data Service enforces data integrity to ensure the accuracy and consistency of data. Fixity in terms of digital preservation refers to the property of a digital file being fixed, or unchanged. Fixity checking is the process of verifying that a digital object has not been altered or corrupted (Wikipedia, 2015). The ICSSR Data Service runs fixity checks on deposited datasets to ensure that the data and the documentation associated with those has not been altered and intact en-route between two points of time or events while transferring, conversion or data compression, etc. and match in terms of special

encodings of variables, variable names, values, value labels, and column number, width and digital signature, etc. Information created by these fixity checks provides evidence for the integrity and authenticity of the digital objects.

5.6. Withdrawal of Data and Succession Plans

5.6.1. Conditions for Withdrawal of Data

The datasets deposited into the repository may be withdrawn or removed from the repository in later phase by the ICSSR Data Service following the request from the data owner or depositor because of the following issues:

- ☞ Copyright violation;
- ☞ Legal requirements and proven legal violations;
- ☞ National security;
- ☞ Falsified and unfair research;
- ☞ Confidentiality concerns, etc.

References

Inter-university Consortium for Political and Social Research (ICPSR). (2012). Guide to Social Science Data Preparation and Archiving: Best Practice Throughout the Data Life Cycle (5th ed.). Ann Arbor, MI. http://www.icpsr.umich.edu/files/deposit/dataprep.pdf

Berlin Declaration. (2003). Open Access to Knowledge in the Sciences and Humanities. http://oa.mpg.de/lang/en-uk/berlin-prozess/berliner-erklarung/

Consortium of European Social Science Data Archives (CESSDA). http://cessda.net/CESSDA-Training

Data Archiving and Networked Services (DANS). (2015). Preservation Policy (Version 1.1). http://www.dans.knaw.nl/en/about/organisation-and-policy/legal-information/dans-preservation-policy-uk.pdf

Elizabeth (Libbie) Stephenson. (2011). Social Science Data Archive Policy on Acquisitions and Archiving (Revised version). http://dataarchives.ss.ucla.edu/index/home_archivepolicies.htm

Engineering and Physical Sciences Research Council (EPSRC). (2015). Policy framework on research data. https://www.epsrc.ac.uk/about/standards/researchdata/scope/

Green, A., Mcdonald, S. & Rice, R. (2009). Policy-making for Research Data in Repositories: A Guide (Ver. 1.2). www.disc-uk.org/docs/guide.pdf

National Informatics Centre (NIC), Government of India. (2014). Implementation Guidelines for National Data Sharing and Accessibility Policy (NDSAP) (Ver. 2.2). https://data.gov.in/sites/default/files/NDSAP_Implementation_Guidelines_2.2.pdf

Office of Science and Technology Policy (OSTP). (2013). Increasing Access to the Results of Federally Funded Scientific Research. Washington D.C., USA. https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

Organisation for Economic Co-operation and Development (OECD). (2007). Principles and Guidelines for Access to Research Data from Public Funding. http://www.oecd.org/sti/sci-tech/38500813.pdf

Research information network (RIN). (2008). Stewardship of digital research data: a framework of principles and guidelines. http://www.rin.ac.uk/system/files/attachments/Stewardship-data-guidelines.pdf

The Royal Society. (2012). Science as an open enterprise. London, UK. https://royalsociety.org/~/media/policy/projects/sape/2012-06-20-saoe.pdf

UK Data Archive. Research Data Lifecycle. http://www.data-archive.ac.uk/create-manage/life-cycle

UK Data Service. http://ukdataservice.ac.uk/

Whyte, A. & Tedds, J. (2011). Making the Case for Research Data Management. Digital Curation Centre. Edinburgh, UK. http://www.dcc.ac.uk/resources/briefing-papers

https://en.wikipedia.org/wiki/File_Fixity