



[HTTP://WWW.ICSSRDATA SERVICE.IN/](http://www.icssrdataservice.in/)

ICSSR Data Service

Indian Social Science Data Repository

Data Deposit and Access Policy Guideline



Indian Council of Social Science Research

Data Deposit and Access Policy Guideline

Contents:

1.	Background	1
2.	Data Acquisition	1
2.1	Content Coverage	
2.1.1	Data Scope and Language	
2.1.2	Kinds of Research Data	
2.1.3	Data Selection for Inclusion	
2.1.4	Version Control	
2.1.5	Data File Formats	
2.1.6	Volume and Size Limitations	
3	Metadata	3
3.1	Access to Metadata	
3.2	Reuse of Metadata	
3.3	Metadata Types and Sources	
3.4	Metadata Schemas	
4	Data Submission	4
4.1	Eligible Data Depositors	
4.2	Moderation by Data Repository	
4.3	Data Quality Requirements	
4.3.1	Responsibility	
4.3.2	Quality Assessment	
4.4	Confidentiality and Disclosure	
4.5	Embargo Period	
4.6	Rights and Ownership	
5	Conditions for Withdrawal of Data	7
6	Data Access	7
6.1	Coverage of Access Policy	
6.1.1	Users Community	
6.1.2	Access to Datasets	
6.1.2.1	Registration	
6.1.2.2	Search and Access	
6.1.3	Use and Reuse of Data Objects	
6.1.3.1	Use and Reuse	
6.1.3.2	Citation	
6.1.3.3	Submission of Publications	
6.1.3.4	Tracking Users and Usage Statistics	
7	Standing Committee	10

1. Background

The ICSSR Data Service has been set-up as a portal for the social science research community and generators of datasets in India to deposit, use, reuse and analyse data with the objective of strengthening research and policy analysis capability. The service is a culmination of signing of a 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 a “Social Science Data Repository” and hosting of NSS and ASI datasets in the repository. The Data Service will also service additional data sets as provided in this Guideline. The INFLIBNET Centre is executing the project of setting-up of ICSSR Data Service with funding from ICSSR.

This policy guideline is meant for individual researchers, institutions and others to deposit and share their research data, access and use data for social science research in India for academic, non-commercial purposes through ICSSR Data Service. This guideline is outlined below.

2. Data Acquisition

2.1. Content Coverage

2.1.1. Data Scope and Language

The scope of ICSSR Data Service will be the data on social science and other related subject domains in a computer processable format. All data and data-related descriptions will be in English till capabilities are developed for processing data in other languages.

2.1.2. Kinds of Research Data

The data repository will include social and statistical data of various national-level surveys on industries, education, employment and un-employment, household consumer expenditure, enterprise, landholdings, administrative data, compiled data in open tabular format like CSV, EXCEL etc. Also, it will have quantitative and qualitative scientific research data generated by various social science institutions, research services and others. Supplementary information or materials associated with the research will be 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.

2.1.3. Data Selection for Inclusion

The ICSSR Data Service will determine the status of data to qualify for inclusion in the data repository on the basis of their:

- ☞ Demonstrated importance, relevance, completeness, availability of metadata;

- ☞ Quality of data based on agency reputation, research & public policy outcomes, comparability, timeliness; and
- ☞ Easy access conditions to facilitate faster acquisition procedures.

It considers mainly:

- ☞ 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 Service will primarily hold 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. acquired from Ministry of Statistics and Programme Implementation (MoSPI).

In addition, the scope of ICSSR Data Service would be expanded to host datasets from the ICSSR and 27 organizations under its purview, other Ministries, Govt. departments, universities, colleges and other institutions of higher learning, NGOs, individual students, researchers and faculty.

2.1.4. 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 will consider the followings 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 will always keep 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.

2.1.5. Data File Formats

The Data Service will support varieties of preferred data formats for hosting 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 will have the liberty to export / download, convert and use the datasets using different statistical software packages according to their requirements.

2.1.6. 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 Service will reserve the right to refuse or accept datasets or accompanied materials. It 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.

3. Metadata

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.

3.1. Access to Metadata

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

3.2. 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.

3.3. 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.

3.4. 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.

4. Data Submission

4.1. Eligible Data Depositors

- ☞ **MoSPI:** Ministry of Statistical Planning and Implementation, Government of India;
- ☞ **ICSSR:** Indian Council of Social Science Research (ICSSR) and its 27 constituent research institutions 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 ICSSR Data Service policies. These contributors may be individuals /

independent researchers, government organizations, research based 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 supporting documents in appropriate format(s) into the repository along with the filled and signed copies of prescribed [Data Deposit Form](#) (Appendix I) and [Licence Agreement Form](#) (Appendix II). Data owners also have the option to send their data offline on CDs / DVDs using postal or Courier Services to the “Data Collection Team, ICSSR Data Service” or sent their data as attachment to the e-mail or share it on Google Drive or Drop Box..

4.2. Moderation by Data 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 and subsequently approved by the Standing Committee. Datasets proposed for deposit would be 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 ICSSR Data Service adds additional keywords and persistent identifiers’ to each datasets deposited into it.

4.3. Data Quality Requirements

4.3.1. 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 ICSSR 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 ICSSR Data Service.

4.3.2. Quality Assessment

It is the sole discretion of the ICSSR Data Service to make decisions about whether to accept a dataset or not. The ICSSR 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 or that has to be created by the ICSSR Data Service. 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.

4.4. 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 will undergo 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.

4.5. 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.

4.6. 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. It is 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. 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 because of the following issues, if found with the deposited data.

- ☞ Copyright violation;
- ☞ Legal requirements and proven legal violations;

- ☞ National security;
- ☞ Falsified and unfair research;

- ☞ Confidentiality concerns, etc.

However, the data owner or depositor may also request to the ICSSR Data Service for withdrawal / removal of his / her deposited data from the repository in unavoidable and desired conditions.

6. Data Access

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).”

6.1. Coverage of Access Policy

6.1.1. Users Community

- ☞ Students: Graduate / Post Graduate students from institutions of higher education, and research institutes in India;
- ☞ Research Scholars: M. Phil. / Ph.D. / Post. Doc. research scholars from Indian academic and research institutes;
- ☞ Academia, faculty members and social scientists from Indian academic and research institutes;
- ☞ Social science research institutions supported by ICSSR;
- ☞ Government ministries, departments, offices, and other research services affiliated to and working under Government of India.
- ☞ Other users i.e. independent researchers, research based NGOs, etc.

- ☞ Foreign Users: Students, research scholars, scientists, and faculty members from abroad institutions with which ICSSR has bilateral understandings / agreements on sharing of resources subject to the condition that similar facilities will be reciprocated by such institutions with respect to resources held by them.

6.1.2. Access to Datasets

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 will facilitate access to each datasets deposited in the repository in accordance with the request and conditions specified by the data owners at the time of depositing dataset and signing license agreement, or any other considerations deemed important by the ICSSR Data Service itself. 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. User can freely download these data directly from the repository.
- ☞ **Safeguarded Data:** Data can be access only to the registered users only after a prescribed online 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.

"However, in context of MoSPI data, dissemination to private agencies, foreign students / institutions and official use will continue to be done by MoSPI".

6.1.2.1. 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 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 "Terms of Use" statement available online 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.

6.1.2.2. 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.

6.1.3. Use and Reuse of Data Objects

6.1.3.1. 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. Individuals or organizations responsible for misuse of the datasets might be debarred from using the ICSSR Data Service. In addition, it may also lead to legal action against such individuals or organizations.

6.1.3.2. 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 users of data from the repository should provide correct citation and must provide acknowledgement to the ICSSR Data Service and the data provider for data used.

User may refer to: [How to Cite Data?](#) for further details.

6.1.3.3. Submission of Publications

Users of the datasets from the ICSSR Data Service shall acknowledge the use of datasets in their publications emanating from the use of data and submit a copy such publications to the ICSSR Data Service.

6.1.3.4. 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.

7. Standing Committee

The ICSSR will constitute a Standing Committee to oversee the implementation of this guideline.

For further information please contact

Dr. Upendra Choudhury

Member-Secretary
Indian Council of Social Science Research
Aruna Asaf Ali Marg
New Delhi - 110067
E-mail: ms@icssr.org

Dr. Jagdish Arora

Director
Information and Library Network Centre
An IUC of UGC
Infocity, Gandhinagar -382007, Gujarat
E-mail: director@inflibnet.ac.in



Indian Council of Social Science Research

JNU Institutional Area, Aruna Asaf Ali Marg
New Delhi - 110067 (INDIA)

For more information : <http://www.icssrdataservice.in/>.