Irish Social Science Data Archive

File Format Policy

The Irish Social Science Data Archive (ISSDA) is Ireland's centre for quantitative data acquisition, preservation, and dissemination. Based at UCD Library, its mission is to ensure wide access to quantitative datasets in the social sciences, and to advance the promotion of international comparative studies of the Irish economy and Irish society.

Introduction

ISSDA's preferred and acceptable formats are widely used by our Designated Community[1]. ISSDA's Designated Community consists of higher education institutions (HEIs) and research-performing organisations (RPOs) across multiple sectors and jurisdictions with broad relevance to the Social Sciences and Public Health.

Preferred formats are formats most suitable for sharing, reuse and preservation, while acceptable formats are less suitable for data preservation but are widely used within the social sciences community. We recommend that Data Depositors deposit data in preferred or acceptable formats. Other formats may still be considered by ISSDA in discussion with the Depositor.

ISSDA recommends that multiple formats are deposited, where possible, to allow as broad a range of researchers to access and use the data as possible.

Dataverse, the open source software utilised by ISSDA for its data repository, automatically transforms SPSS and other statistical software into plain text, TAB-delimited files. The metadata is stored separately as an XML file. These can be read together into SPSS and other statistical applications. This allows as broad a range of researchers as possible to access the data without the need for proprietary software.

---

[1] The term 'Designated Community' is an OAIS terms and refers to an identified group of potential data consumers who should be able to understand a particular set of information from an archive.

This also supports the long term archiving and preservation of data as ISSDA recommends Tab-delimited file (.tab) with setup file (for SPSS, Stata, SAS, etc.) for this purpose.

If the Data Depositor is unable to deposit data in the preferred formats, an effort will be made to derive files in preferred formats where the source file format can be reliably converted.

ISSDA Preferred and Acceptable Formats

Type of data	Preferred formats	Acceptable formats
Quantitative Datasets (statistical file formats)	• Tab or comma-delimited file (.tab, csv) with setup file (for SPSS, Stata, SAS, etc.)	• SPSS (.por, .sav) • Stata (.dta) • SAS (.7bdat; .sd2; .tpt) • Tab- or comma-delimited text documents (e.g., .csv, .tab, .tsv), without import syntax/script.
Documentation (text documents)	• PDF/A (.pdf)	• PDF (.pdf) • ODT (.odt) • MS Word (.doc, .docx) • RTF (.rtf)
Metadata	• DDI (.xml)	

As a result of our close relationship with our Designated Community we are aware of the importance of availability of data in these formats to them. We are also aware that the use of R for statistical analysis is becoming increasingly popular. Our preferred formats can be read into R and so largely meet the current needs of our Designated Community using R.

Documentation is converted to PDF for dissemination purposes, where possible.

In preparing these preferred and acceptable formats we have consulted the recommendations of the following organisations: Digital Archiving and Networked Services (DANS)[2], Leibniz-Institut für Sozialwissenschaften (GESIS)[3], Inter-university Consortium for Political and Social Research (ICPSR)[4] and the UK Data Archive[5].

Review

This Policy will be reviewed regularly in light of any changes in technology or the requirements of our Designated Community.

---

[2] https://dans.knaw.nl/en/file-formats/
[3] https://www.gesis.org/en/data-services/about-the-data-services/standards-and-workflows-data-services/preparation-of-data-transfer-pre-ingest
[4] https://www.icpsr.umich.edu/icpsrweb/content/datamanagement/policies/colldev.html
[5] https://ukdataservice.ac.uk/learning-hub/research-data-management/format-your-data/recommended-formats/