

***Planes de Gestión de Datos en
Humanidades y Ciencias
Sociales***



Universitat, coneixement i llibertat
Universidad, conocimiento y libertad
University, knowledge and freedom



Características de los datos de Humanidades/ Ciencias Sociales

Gran diversidad de datos primarios para la investigación



Yacimientos arqueológicos Artefactos
Textos
Mapas
Manuscritos
Fotografías

Características de los datos de Humanidades/ Ciencias Sociales

Gran diversidad en métodos de recolección, análisis y difusión



Se requiere un enfoque disciplinar

Heterogeneidad de estándares y protocolos

Repositorios temáticos, infraestructuras de redes y agregadores, bibliotecas digitales, revistas, proyectos de Humanidades Digitales ..

Características de los datos de Humanidades/ Ciencias Sociales

Tratamiento de datos personales/sensibles y licencias específicas



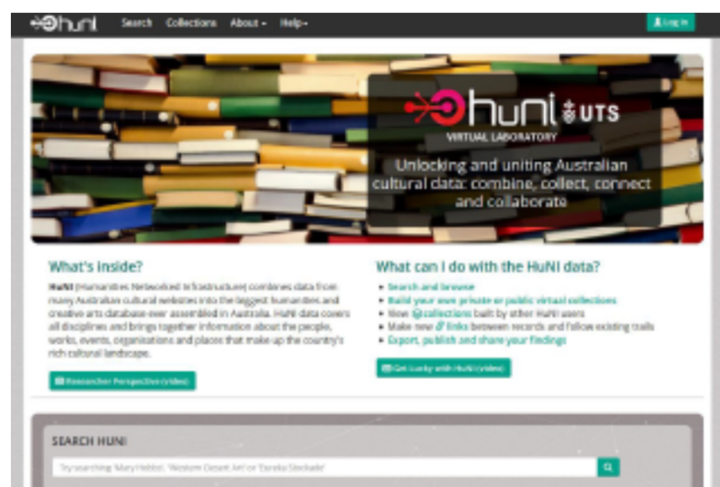
Muchos investigadores recelan de usos comerciales y/o de obras derivadas de sus datos
Necesidad de conocer licencias y estándares específicos para Humanidades/Ciencias Sociales

¿Por qué compartir los datos de investigación?

- Promueve la investigación y el debate
- Promueve la innovación y nuevos usos para los datos
- Da paso a nuevos tipos de colaboraciones entre los creadores y los usuarios de los datos
- Aumenta la transparencia y la rendición de cuentas de los proyectos
- Permite el análisis de las conclusiones de las investigaciones
- Anima a mejorar y validar los métodos de investigación
- Reduce el coste de la duplicación en la recolección de datos
- Aumenta el impacto y la visibilidad de la investigación
- Da crédito a los creadores de los datos
- Proporciona nuevos recursos educativos/de investigación

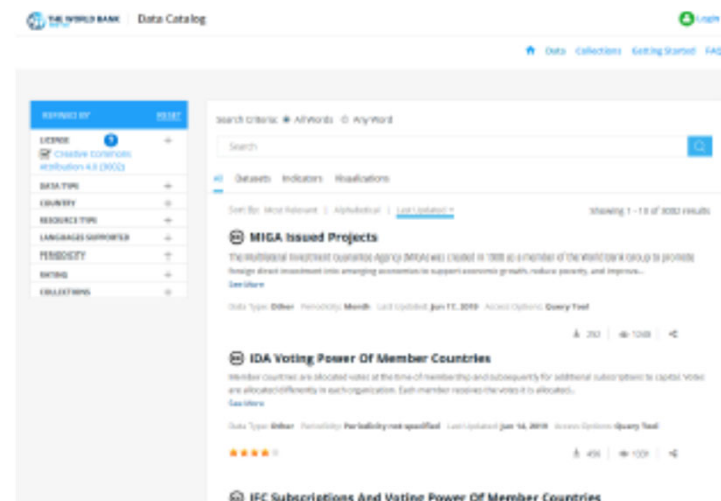
Ejemplos

Humanidades



<https://huni.net.au/#/search>

Ciencias Sociales



<https://data.worldbank.org/>

El papel de la Comisión Europea en impulsar la planificación y la difusión de los datos FAIR



Recomendaciones para los Investigadores

- Define las responsabilidades para la gestión de datos entre los miembros del equipo en el plan de gestión de datos
- Acuerda el uso de estándares comunes para recoger, almacenar y describir datos
- Asegúrate de que todos los datos resultantes del proyecto se compartan en una plataforma común y segura entre todos los miembros del equipo
- Documenta la metodología usada en la recogida de datos y las decisiones de una manera sistemática, por ejemplo, usando plantillas comunes
- Usa estándares comunes para denominar a los ficheros, formatos, estructura de los contenidos

Recomendaciones para los Investigadores

- Pon por escrito cómo se gestionan los datos y se comparten entre los miembros del equipo
- Establece qué datos son de uso interno para los miembros del equipo y cuáles se difundirán públicamente
- Deposita los datos en un repositorio que permite el acceso a los revisores de las publicaciones resultantes del proyecto
- Elige licencias de uso estándares para los datos resultantes del proyecto
- Asigna un identificador persistente (DOI, Handle, PURL...) a los datos resultantes para que puedan ser encontrados, reutilizados y citados
- Publica información (METADATOS) sobre los datos resultantes en repositorios

Las preguntas que debes hacerte para que tus datos sean FAIR

- ¿Tiene el dataset un identificador persistente (handle, DOI)?
- ¿Hay documentación/metadatos que permitan entender los datos adecuadamente?
- ¿Los metadatos están accesibles?
- ¿Se ha aplicado al dataset una licencia de uso? ¿Es una licencia estándar? ¿Impone algún tipo de restricciones? ¿Permite explícitamente la reutilización?
- ¿Los ficheros del dataset se encuentran en formatos abiertos o en formatos propietarios ampliamente soportados?
- ¿Están el dataset/metadatos codificados según algún estándar global?
- ¿Está el dataset enlazado a otros datasets u otros resultados de investigación? ¿Cómo?

Generadores de Planes de Gestión de Datos

The screenshot shows the DMP ONLINE website interface. At the top, there is a navigation bar with the logo 'DMP ONLINE' and links for 'Home', 'FAQ/Help', 'Funder requirements', and 'Help'. The main content area features a 'Welcome' message, a description of the service, and statistics: 17,622 Users, 203 Organisations, 23,083 Plans, and 88 Countries. A sign-in form is visible on the right side, including fields for Email and Password, a 'Forgot password?' link, a 'Remember me!' checkbox, and buttons for 'Sign in' and 'Create account'. A footer note mentions that some funders mandate the use of DMP ONLINE, while others just recommend it as a useful option.

Herramienta de facto para preparar planes para proyectos :

<https://dmponline.dcc.ac.uk/>

https://dmponline.dcc.ac.uk/public_plans

PARTHENOS: DMP para Humanidades

DMP component	Issues to be addressed	Guidance
1. Data summary	State the purpose of the data collection/generation	Please, include a brief description of the reason for collecting/generating data
	Explain the relation to the objectives of the project	<input type="checkbox"/> Data availability <input type="checkbox"/> Data reuse <input type="checkbox"/> Data interoperability <input type="checkbox"/> Other, please specify
	Specify the types and formats of data generated/collected	<input type="checkbox"/> All data/formats are collected <input type="checkbox"/> Only open formats are collected <input type="checkbox"/> List data formats: i.e. XML, RDF, TEI... <input type="checkbox"/> Include link to preferred data format document <input type="checkbox"/> Select format from the suggested lists: <ul style="list-style-type: none"> <input type="checkbox"/> CLARIN <input type="checkbox"/> Meertens Institute <input type="checkbox"/> CNRS <input type="checkbox"/> KNAW-DANS <input type="checkbox"/> GAMS <input type="checkbox"/> ADS <input type="checkbox"/> Other, please specify
	State the expected size of the data	<input type="checkbox"/> Number of files, please specify <input type="checkbox"/> Number of digital objects, please specify <input type="checkbox"/> Not available
	Specify the granularity of the collected data	<input type="checkbox"/> Single items (i.e. one

DMP component	Issues to be addressed	Guidance
	Outline the data utility: to whom will it be useful	Please, list possible stakeholders reusing your data
	Describe your strategy of data exploitation	Please specify if you: <ul style="list-style-type: none"> <input type="checkbox"/> Plan agreements with other institutions <input type="checkbox"/> Use policy on data reuse <input type="checkbox"/> Use license of use, specify which <input type="checkbox"/> Other, please specify
	2. FAIR Data 2.1. Making data findable, including provisions for metadata	Specify standards for metadata creation (if any). If there are no standards in your discipline describe what type of metadata will be created and how. Please, select from the list: <ul style="list-style-type: none"> <input type="checkbox"/> ACDM <input type="checkbox"/> CAIRARE <input type="checkbox"/> CCR <input type="checkbox"/> CIDOC CRM <input type="checkbox"/> CMDCI <input type="checkbox"/> DC <input type="checkbox"/> DDI <input type="checkbox"/> Europeana <input type="checkbox"/> ICCD <input type="checkbox"/> MIDAS <input type="checkbox"/> OAI-ORE <input type="checkbox"/> Other, please specify
	Specify if metadata are updated during the project and/or once the data are archived	<input type="checkbox"/> Yes, automatically <input type="checkbox"/> Yes, manually <input type="checkbox"/> No <input type="checkbox"/> I don't know
	Describe the mechanisms used to identify digital resources	<input type="checkbox"/> Digital Object Identifiers <input type="checkbox"/> Aleph System number <input type="checkbox"/> Data Cite DOI <input type="checkbox"/> Digital resources ID

https://www.rd-alliance.org/sites/default/files/attachment/PARTHENOS%20DMP_draft.pdf

Elementos comunes en todos los Planes de Gestión de datos

Theme	DCC & UC3 Guidance
DATA DESCRIPTION	<ul style="list-style-type: none"> • Give a summary of the data you will collect or create, noting the content, coverage and data type, e.g., tabular data, survey data, experimental measurements, models, software, audiovisual data, physical samples, etc. • Consider how your data could complement and integrate with existing data, or whether there are any existing data or methods that you could reuse. • Indicate which data are of long-term value and should be shared and/or preserved. • If purchasing or reusing existing data, explain how issues such as copyright and IPR have been addressed. You should aim to minimise any restrictions on the reuse (and subsequent sharing) of third-party data.
DATA FORMAT	<ul style="list-style-type: none"> • Clearly note what format(s) your data will be in, e.g., plain text (.txt), comma-separated values (.csv), geo-referenced TIFF (.tif, .tiff). • Explain why you have chosen certain formats. Decisions may be based on staff expertise, a preference for open formats, the standards accepted by data centres or widespread usage within a given community. • Using standardised, interchangeable or open formats ensures the long-term usability of data; these are recommended for sharing and archiving. • See UK Data Service guidance on recommended formats or DataONE Best Practices for file formats.
DATA VOLUME	<ul style="list-style-type: none"> • Note what volume of data you will create in MB/GB/TB. Indicate the proportions of raw data, processed data, and other secondary outputs (e.g., reports). • Consider the implications of data volumes in terms of storage, access and preservation. Do you need to include additional costs? • Consider whether the scale of the data will pose challenges when sharing or transferring data between sites; if so, how will you address these challenges?

- Roles y responsabilidades
- Descripción, volumen, estructura, estándares, formatos de los datos
- Recogida y procesamiento de datos
- Esquema de Metadatos
- Propiedad intelectual y licencias de uso
- Consideraciones éticas
- Acceso, publicación y reutilización de los datos
- Selección del repositorio
- Seguridad y almacenamiento
- Preservación a largo plazo
- Presupuesto

Formatos yesquemas de metadatos estándar

2.5	Spreadsheets	<ul style="list-style-type: none"> • ODS (.ods) • CSV (.csv) 	<ul style="list-style-type: none"> • MS Excel (.xls, .xlsx) • PDF/A (.pdf) • OOXML (.docx, .docm)
2.6	Databases	<ul style="list-style-type: none"> • SQL (.sql) • SIARD (.siard) • DB tables (.csv) 	<ul style="list-style-type: none"> • MS Access (.mdb, .accdb), (v. 2000 or later) • dBase (.dbf) (v. 7 or later) • HDF5 (.hdf5, .he5, .h5)
2.7	Statistical data	<ul style="list-style-type: none"> • SPSS Portable (.por) • SPSS (.sav) • STATA (.dta) • DDI (.xml) • data (.csv) + setup (.txt) 	<ul style="list-style-type: none"> • SAS (.7bdat; .sd2; .tpt) • R^(*)
2.8	Raster Images	<ul style="list-style-type: none"> • JPEG (.jpg, .jpeg) • TIFF (.tif, .tiff) • PNG (.png) • JPEG 2000 (.jp2) 	<ul style="list-style-type: none"> • DICOM (.dcm) (by mutual agreement)
2.9	Images (vector)	<ul style="list-style-type: none"> • SVG (.svg) 	<ul style="list-style-type: none"> • Illustrator (.ai) • EPS (.eps)
2.10	Audio	<ul style="list-style-type: none"> • WAVE; BWF (.wav) • FLAC (.flac) 	<ul style="list-style-type: none"> • AIFF (.aif, .aiff) • MP3 (.mp3) • AAC (.aac, .m4a)
2.11	Video	<ul style="list-style-type: none"> • MPEG-2 (.mpg, .mpeg, ...) • MPEG-4 H.264 (.mp4) • Lossless AVI (.avi) • QuickTime (.mov) 	<ul style="list-style-type: none"> • MKV (.mkv)
2.12	Computer Aided Design (CAD)	<ul style="list-style-type: none"> • AutoCAD DXF v. R12 (.dxf) 	<ul style="list-style-type: none"> • AutoCAD, other versions (.dwg, .dxf) • ESRI Shapefiles (.shp &

Formatos usados frecuentemente por la comunidad científica Usan especificaciones abiertas Independientes de software/empresas específicas

Formatos yesquemas de metadatos estándar

2.5	Spreadsheets	<ul style="list-style-type: none"> • ODS (.ods) • CSV (.csv) 	<ul style="list-style-type: none"> • MS Excel (.xls, .xlsx) • PDF/A (.pdf) • OOXML (.docx, .docm)
2.6	Databases	<ul style="list-style-type: none"> • SQL (.sql) • SIARD (.siard) • DB tables (.csv) 	<ul style="list-style-type: none"> • MS Access (.mdb, .accdb), (v. 2000 or later) • dBase (.dbf) (v. 7 or later) • HDF5 (.hdf5, .he5, .h5)
2.7	Statistical data	<ul style="list-style-type: none"> • SPSS Portable (.por) • SPSS (.sav) • STATA (.dta) • DDI (.xml) • data (.csv) + setup (.txt) 	<ul style="list-style-type: none"> • SAS (.7bdat; .sd2; .tpt) • R^(*)
2.8	Raster Images	<ul style="list-style-type: none"> • JPEG (.jpg, .jpeg) • TIFF (.tif, .tiff) • PNG (.png) • JPEG 2000 (.jp2) 	<ul style="list-style-type: none"> • DICOM (.dcm) (by mutual agreement)
2.9	Images (vector)	<ul style="list-style-type: none"> • SVG (.svg) 	<ul style="list-style-type: none"> • Illustrator (.ai) • EPS (.eps)
2.10	Audio	<ul style="list-style-type: none"> • WAVE; BWF (.wav) • FLAC (.flac) 	<ul style="list-style-type: none"> • AIFF (.aif, .aiff) • MP3 (.mp3) • AAC (.aac, .m4a)
2.11	Video	<ul style="list-style-type: none"> • MPEG-2 (.mpg, .mpeg, ...) • MPEG-4 H.264 (.mp4) • Lossless AVI (.avi) • QuickTime (.mov) 	<ul style="list-style-type: none"> • MKV (.mkv)
2.12	Computer Aided Design (CAD)	<ul style="list-style-type: none"> • AutoCAD DXF v. R12 (.dxf) 	<ul style="list-style-type: none"> • AutoCAD, other versions (.dwg, .dxf) • ESRI Shapefiles (.shp &

Iniciativa internacional de voluntarios para crear un directorio de esquemas de metadatos que son estándares internacionales

Formatos yesquemas de metadatos estándar

Open research data and data deposition in the Social Sciences and Humanities domain

The situation with regard to open data in the SH domain, both in terms of infrastructure (depositories) as well as protocols and standards, is rapidly evolving. There are many initiatives, at the national and supra-national levels, that aim to provide researchers with the necessary tools and information.

Characteristic feature of the disciplines that together make up the ERC's SH domain is their variety, in terms of topics, epistemologies, and methodologies. This is reflected also in the data that SH projects produce: quantitative data sets; experimental data; observational data; interviews; archival data; human artefacts; medical and genetic data; and so forth. In addition, the various kinds of data crosscut the disciplinary divisions, as several disciplines produce different kinds of data, depending on the methodologies used.

Also, particular restrictions may apply to making data open depending on the discipline. Data may include copyrighted material, such as literary texts or images, or archival materials to which access is restricted. In other cases, data may include privacy-sensitive material, such as video recordings of parent-child interactions or interviews.

For this reason, it is not possible to provide a single set of guidelines for the entire SH domain. Therefore, this document aims to provide some general and some discipline-specific references that ERC grantees can use to draw up DMPs that are adequate for their discipline and their specific project, and that meet the FAIR principles.

In what follows more information is given on:

- general depositories
- discipline-specific depositories
- metadata and data preparation

General depositories

There are many options available for SH scholars, both general as well as discipline-specific, not-for-profit as well as commercial. The list below mentions a number of well-known depositories for use by social sciences and humanities disciplines, but it is certainly not exhaustive.

An important selection of depositories for SH scholars is provided by CESSDA:

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

CESSDA is a so-called 'ERIC', a European Research Infrastructure Consortium (http://ec.europa.eu/research/infrastructure/index_en.htm?pg=eric), i.e., an international entity established by the European Commission that has national governments or consortia as its members. Currently, CESSDA has 18 members, all of them national agencies that operate on a not-for-profit basis.

Many of the CESSDA depositories also cover (some of) the humanities in addition to the social sciences.