SERA
Seismology and Earthquake Engineering
Research Infrastructure Alliance for Europe

DELIVERABLE

D5.5 Validation of pre-operational access phase to selected DSS datasets and products

Work package	D5.5 Validation of pre-operational access phase to selected DSS datasets and products
Lead	CSIC, UU
Authors	DeFelipe, Irene, CSIC Carbonell, Ramon, CSIC Ivandic, Monika, Uppsala University Roland, Roberts, Uppsala University
Reviewers	-
Approval	Management Board
Status	Final
Dissemination level	Public
Delivery deadline	30.04.2020
Submission date	27.04.2020
Intranet path	DOCUMENTS/DELIVERABLES/ SERA_D5.5_pre-operational access phase to selected DSS.pdf

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 730900.

Table of Contents

Summary

This deliverable is written within the framework of the project “Seismology and Earthquake Engineering Research Infrastructure Alliance for Europe – SERA” (Project no: 730900), funded by the Horizon2020, INFRAIA-01-2016-2017 Programme.
This deliverable aims to present the pre-operational access phase to DSS data and data products and a sucessful prototype of an e-infrastructure for sharing open data, the Institute of Earth Sciences Jaume Almera (ICTJA-CSIC) dataset. It regards for the importance of sharing seismic data, data products and facilities following the FAIR principles. Thus, the ICTJA-CSIC dataset is a multidisciplinary e-infrastructure comprising Solid Earth Science data and data products.

1 Introduction

During the last decades, a large amount of scientific data has been produced by the earthquake engineering and seismology research infrastructures in Europe. Nowadays in Europe, there are mainly two organizations regarding for storing this kind of data, namely ORFEUS[1] (Observatories and Research Facilities for European Seismology) for seismological waveform data and EMSC[2] (European Mediterranean Seismological Centre) for earthquake source parameters. In addition to this, the ESFRI[3] (European Strategic Forum for Research Infrastructures) initiative and the EPOS[4] project (European Plate Observing System), provided a larger framework for the integration of all Solid Earth Science data into a single European e-infrastructure.

Despite all the effort carried out by these European organizations, there is still a need to increase the interaction between the earthquake engineering and seismology communities. Special interest is set on the integration of the most important data e-infrastructures and related informatics services in Europe. An efficient use of the resources and know-how sharing data along with the added value of bringing together the data sources and data exchange services of the two communities, will be a step forward in the provision of tools and knowledge for the benefit of researchers and industry users. Such e-tools are an emerging trend to manage and share results following the FAIR principles of findable, accessible, interoperable and reusable data.

Among the different projects providing access to data, stands out SERA[5]. Its aim is to reduce the risk posed by natural and anthropogenic earthquakes based on innovative research and development projects. SERA will significantly improve the access to data, services and research infrastructures and will spread the scientific knowledge of earthquake engineering and seismology by making research data more accessible. The availability and accessibility of these data, therefore, is of foremost importance for the society, including scientists, decision-makers and finally, the general public. Figure 1 shows the organizational structure of SERA, where a coordinator is responsible of the management board, communications, the SERA office and the General Assembly. The Management board delegates in the work package leaders the join research activities, data access and networking. Among all the work packages, the WP5 comprises a networking for Deep Seismic Sounding (DSS) data and products. This work package will therefore create a database including DSS data acquired in the EU since the late 70's.

[1] https://www.orfeus-eu.org/
[2] https://www.emsc-csem.org/#2
[3] https://ec.europa.eu/info/research-and-innovation/strategy/european-research-infrastructures/esfri_en
[4] https://www.epos-ip.org/
[5] http://www.sera-eu.org/en/home/

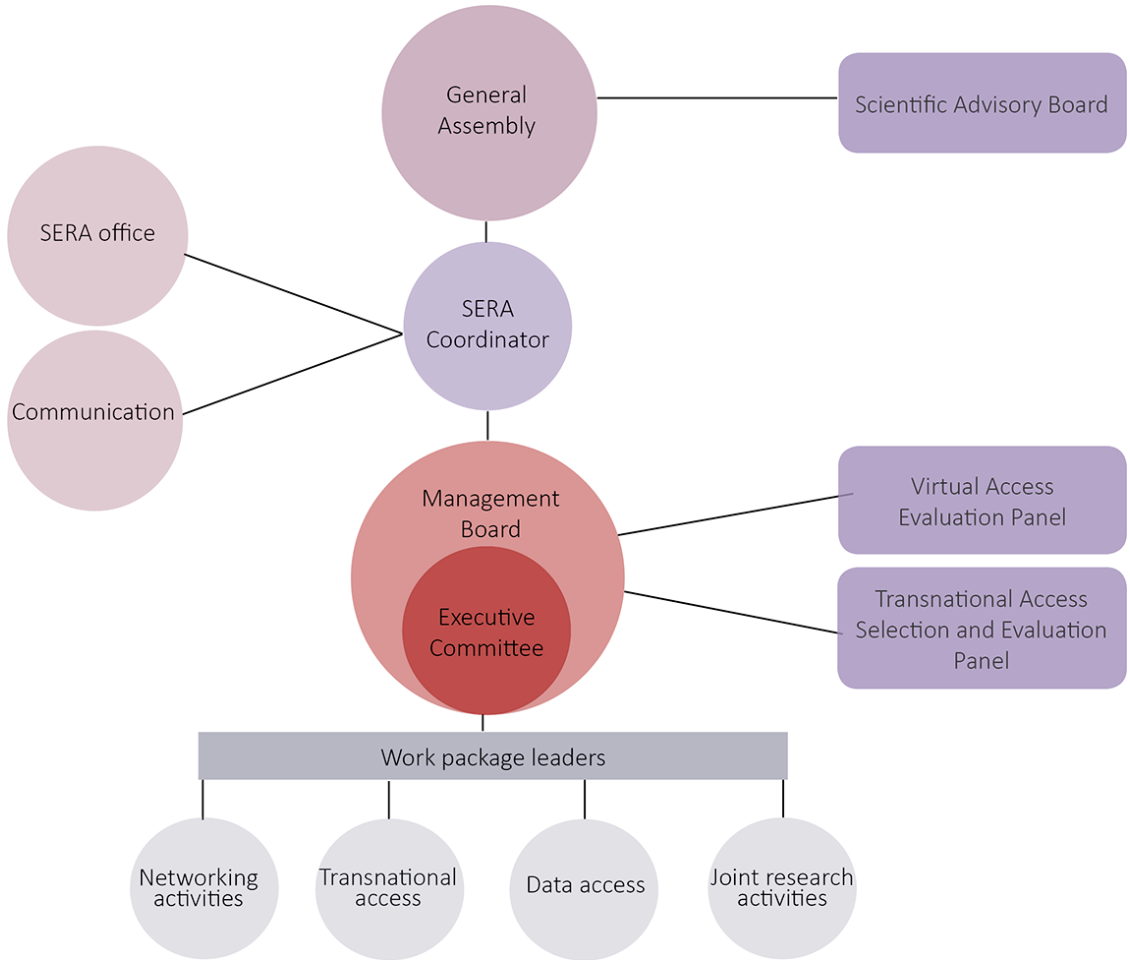


Figure 1: organization structure of SERA (http://www.sera-eu.org/en/about/organisation/).

Controlled source seismic data acquisition experiments have produced a vast amount of DSS data since its development in the early 1950 (e.g., Mooney et al., 2002). The huge amount of DSS data generated provides a valuable information on the structure and nature of the crust and lithosphere, which constitutes a fundamental and robust basis for research within Solid Earth Science. These datasets are unique and constitute the output of an enormous and very expensive scientific effort. Their uniqueness and recognized value evidences the need for their preservation, for this reason the EU is promoting science open access by means of networks, infrastructures and international projects. In addition, scientists are gaining awareness of the importance of sharing the raw data and products open access for future research as well as to reuse old and legacy data. Furthermore, the new developments in processing and achievements in imaging technologies generates new possibilities for these old datasets.

In this deliverable we present the open access e-infrastructure of the Institute of Earth Sciences Jaume Almera (ICTJA[6]) of DSS data and data products. In Spain, the Spanish National Research Council (CSIC) hosts an online platform, DIGITAL.CSIC[7], where scientific data are accessible for the general public (Bernal, 2011; DeFelipe et al., 2019c). Thus, our dataset is the result of a close collaboration with the repository DIGITAL.CSIC. This dataset is available in https://digital.csic.es/handle/10261/101879, following the principles of SERA of open science data and it is organized ensuring an

[6] www.ictja-csic.es
[7] https://digital.csic.es/

intercommunication between different departments and research institutions and promoting networking activities.

2 Operational access phase for DSS dataset

A successful e-infrastructure to access DSS datasets and products aims to achieve all the principles of FAIR data: findable, accessible, interoperable and reusable (Figure 2). The first step in (re)using the data is to find them by means of a globally unique and persistent identifier, such as DOI or handle. Once the users find the data, they need to know how to access the data, possibly including an authentication, and ideally by direct download from the e-infrastructure. Data also needs to be integrated with related data in order to be interoperable with applications or workflows of processing raw data. Finally, the last goal of FAIR is to promote the reuse of the data. For that purpose, data has to be accompanied with metadata, a well-described file that allow to replicate data.

In order to reach all these principles a pre-operation phase in accessing the DSS data needs to be accomplished.

Figure 2: FAIR data principles.[8]

[8] https://findwise.com/blog/data-that-really-saves-lives-and-possibly-your-organisation/

2.1.1 Preparatory phase and networking

The success of a pre-operational access phase to selected DSS datasets and products depends on several factors at different levels. It ranges from individual scientists and reseach groups, to national institutions and international organizations. The effort and work done so far in the framework of SERA, includes an important part of dissemination of national and international mandates in open data as well as promoting informational talks, newsletters and workshops. The following actions are requiered in order to guarantee the current and future access to DSS datasets and products:

- Provide information about the national and international mandates for open data, thus raising awareness of scientists in sharing their research data.
- Promote data management plans at the beginning of a new research project, as a part of the project proposal.
- Encouraging networking between multidisciplinary and international research groups.
- Collaboration between industry and academia to have a mutual benefit.
- Peer-review publications of databases in specific journals.
- A fluent communication between scientists and the responsible of the information resources for research.
- To introduce and spread the concepts of FAIR data and open science into the society at different levels:
    - Local: sharing information of the different research groups through the research institutes websites[9], twitter account[10], and informational talks (Bernal, 2019).
    - National: meetings such as *Bibliotecas y Archivos del CSIC por la Ciencia Abierta*[11], contributions to Earth Sciences conferences such as *Congreso Geológico de España 2020*[12] and internal informational journals such as the *CSIC Abierto*[13].
    - International: attendance to scientific meetings, workshops, and networking activities (Carbonell et al., 2020; DeFelipe et al., 2019a, b, c, 2020) to have these concepts more and more present in the scientific community.

2.1.2 Current state of the DSS datasets

Following the FAIR principles, data have to be accompanied with a detail information file, namely the metadata. These files have to provide the necessary information for new scientists to reprocess or reuse them. An example of metadata is that of the DIGITAL.CSIC repository in coordination with EPOS. The files accompanying the DSS are:

- README.txt (annex I): a text file summarizing the general information regarding the dataset. This document specifies the list of authors, title, license, citation, abstract, referring articles and a list of the attached files.
- Metadata as an excel file (annex II): the metadata consists of two type of information:
    - Common metadata: this section comprises the title; publication date; DOI; licences and rights; authors with their affiliations, their role in the project and their ORCID; a contact person; abstract and methods description; spatial and temporal coverage of the data; related publications; and funding references.
    - Specific metadata: this part of the metadata regards for the type of processing of the DSS data (raw, field or processed), data type (seismic active or passive), data

[9] http://www.ictja.csic.es/
[10] https://twitter.com/ictja_csic?lang=en
[11] http://jornadas.bibliotecas.csic.es/node/4
[12] https://sge-congreso.net/
[13] http://hdl.handle.net/10261/190962

provenance (digital or analogue acquisition), the data file format (normally SEGY), processed method (reflection or refraction), the receiver type and the source type.

- Description of the attached files (annex III): an additional brief description of all the data that has been uploaded.
- According to the type of data uploaded, additional information can be included. These files will regard for location maps, geometry of the acquisition and other type of images or documents.

3 Current state of the ICTJA-CSIC database

Following the national and international mandates of open access data, the FAIR principles (Figure 2) and the awareness of the scientific community in the importance of open data, scientists of the ICTJA-CSIC are making a constant effort to update its database.

During its 50 years' history, the ICTJA-CSIC has generated numerous data not only in the field of geophysics but also in mineral geochemistry, volcanology, satellite imaging, etc. It has resulted in a successful prototype of seismic data repository in collaboration with DIGITAL.CSIC (https://digital.csic.es/handle/10261/101879, last access March 2020). This dataset has been increased especially since the last two years, thanks to the effort of different national and international entities/infrastructures and the increased awareness of scientists in sharing its data.

The type of data included in this dataset comprise active seismic reflection profiles both of normal incidence and wide angle carried out in the Iberian Peninsula since the 90's. The data covers a wide range of geographical and geological settings, sampling both the Variscan and Alpine orogens offshore (Atlantic Iberian Margin, Bay of Biscay, Valencia Trough…) and onshore (Iberian Massif, Cantabrian Mountains, Betic Cordillera…). Our dataset also covers different scales, regarding for continental and crustal scale experiments to local and shallower experiments in the Iberian Peninsula. Finally, our dataset includes both 2D and 3D experiments. It is a large and rich database resulted from a huge and expensive effort of different scientific groups during decades that needs to be preserved for future studies and promoting re-interpretations. Therefore, this dataset strives to make easily available old data but also to set the basis for future data management of new acquired seismic data.

The breakthrough of this dataset can be envisaged based on the statistics of total visits and downloads. Figure 3 shows two statistics plots generated by the DIGITAL.CSIC options and e-facilities for the ICTJA-CSIC dataset. The statistics based on the total number of visits and downloads show the increased interest in the research carried out in the ICTJA-CSIC. The number of visits has been significantly increased since the first months of 2018 onwards. The tendency presents frequent high excursions with several peaks in visitor numbers (Figure 3a). With respect to the total downloads (Figure 3b), an outstanding increase is observed since October 2019. This month correspond to an active period in the DIGITAL.CSIC repository, when several projects were uploaded. These statistics suggest that the scientific community is responding positively to the increase of the number of datasets, evidencing the importance of sharing seismic raw data to move the scientific community forward.

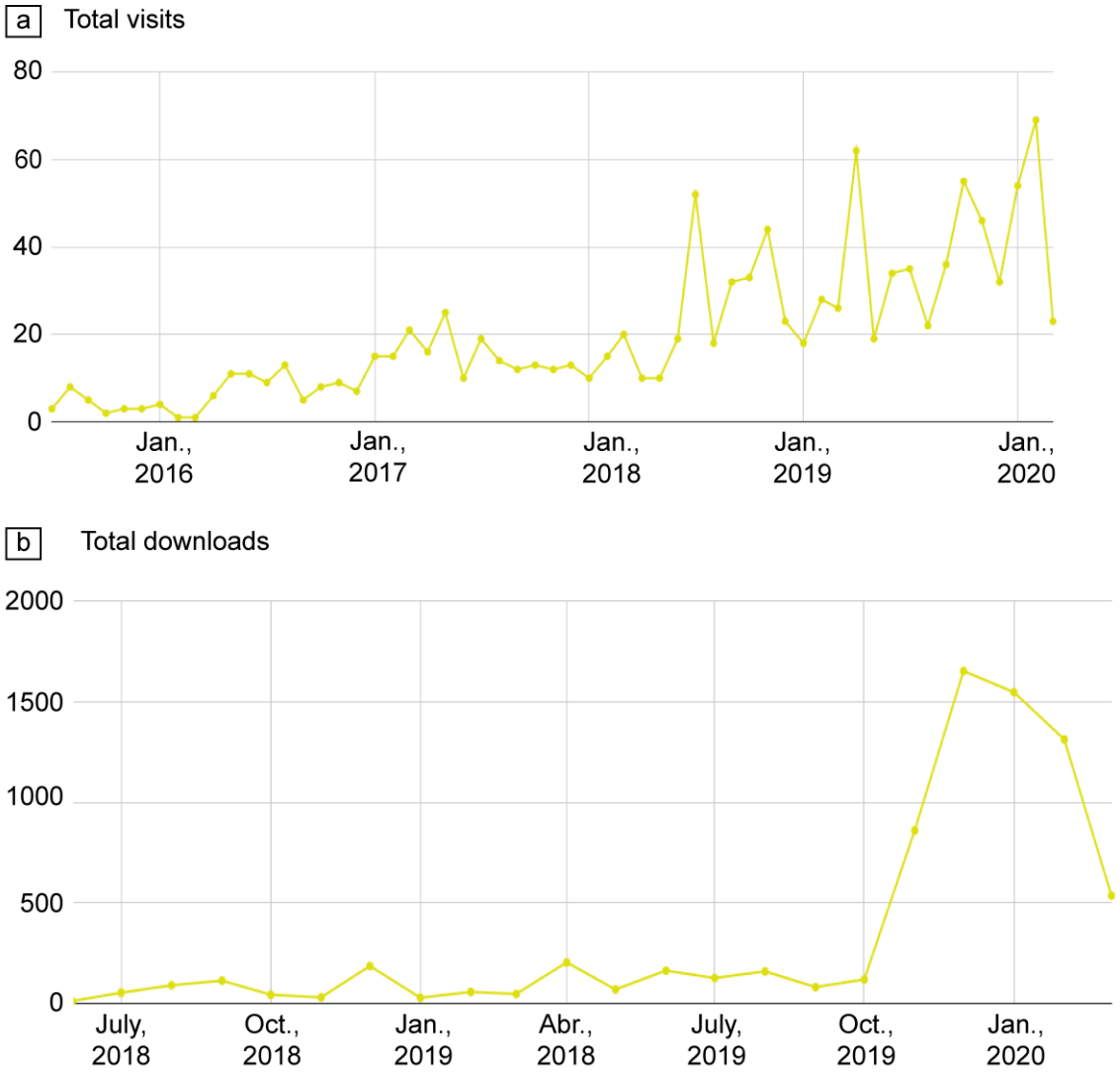


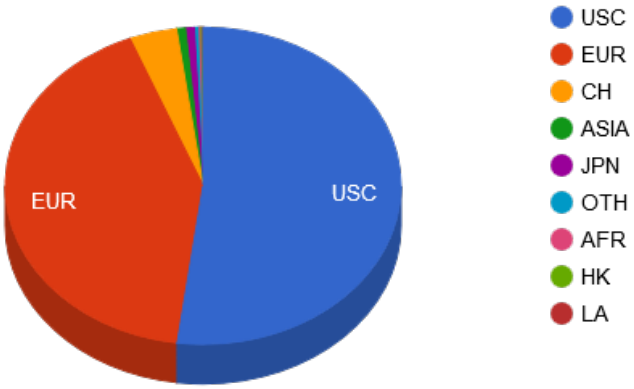
Figure 3: a) total visits to the ICTJA-CSIC dataset, and b) total downloads of data from the same dataset (https://digital.csic.es/handle/10261/101879, last access March 2020).

Our repository provides additional facilities of data mining. Not only it comprises the amount of visits and downloads of any of the projects included in the whole dataset, but also through geographical region, country and city. A summary of this statistical data by region is presented in Figure 4 and Tables 1 and 2 (data for March 2020).

Surprisingly, the regions with more visits to the ICTJA-CSIC dataset are USA and Canada, with a total of 362 visits representing a 52.1%. It is followed by European users that represents the 41.9% (Figure 4a and Table 1). Regarding the total downloads (Figure 4b and Table 2), again, USA and Canada lead the statistics with a total of 749 downloads and representing a 52.3% of the total. It is followed by users in Europe (26.8 %) and in the third place stands Chinese users, representing the 17.1 %.

These data are being constantly updated but they depict the interest of international users in our DSS datasets, proving the effective data management plan and the fruitful dissemination activities carried out so far. It also managed to achieve the FAIR principles, being data findable and accessible for users around the world.

a.



b.

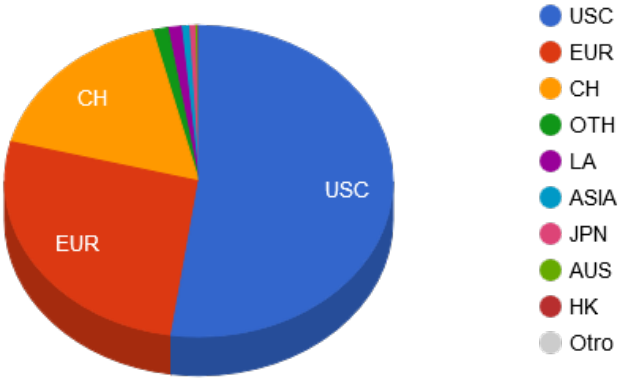


Figure 4: a) total visits and b) total downloads by region. USC: USA and Canada, EUR: Europe, CH: China, OTH: others or not defined, LA: Latinamerica, ASIA: Asia, JPN: Japan, South Korea and Taiwan, AUS: Autralia and New Zeeland, HK: Hong Kong, AFR: Africa (https://digital.csic.es/handle/10261/101879, last access March 2020).

Table 1: total visits by continent (https://digital.csic.es/handle/10261/101879, last access March 2020).

CONTINENT	NUMBER	PERCENTAGE %
USA & CANADA	362	52.1
EUROPE	291	41.9
CHINA	27	3.9
ASIA	5	0.7
JAPAN, S KOREA, TAIWAN	5	0.7
OTHERS	2	0.3
AFRICA	1	0.1
HONG KONG	1	0.1
LATINAMERICA	1	0.1
TOTAL	695	

Table 2: total downloads by continent (https://digital.csic.es/handle/10261/101879, last access March 2020).

CONTINENT	NUMBER	PERCENTAGE %
USA & CANADA	749	52.3
EUROPE	384	26.8
CHINA	245	17.1
OTHERS	18	1.3
LATINAMERICA	16	1.1
ASIA	9	0.6
JAPAN, S KOREA, TAIWAN	7	0.5
HONG KONG	2	0.1
AUTRALIA & NEW ZEALAND	2	0.1
AFRICA	1	
TOTAL	1433	

4 Conclusions

Due to the increase in the volume of DSS data generated, a data management plan is indispensable in order to have data freely available to download for the scientific community, industry and the general public. A successful prototype is presented here, the ICTJA-CSIC database which aims to share new and old seismic data following the FAIR principles of data management: findable, accessible, interoperable and reusable.

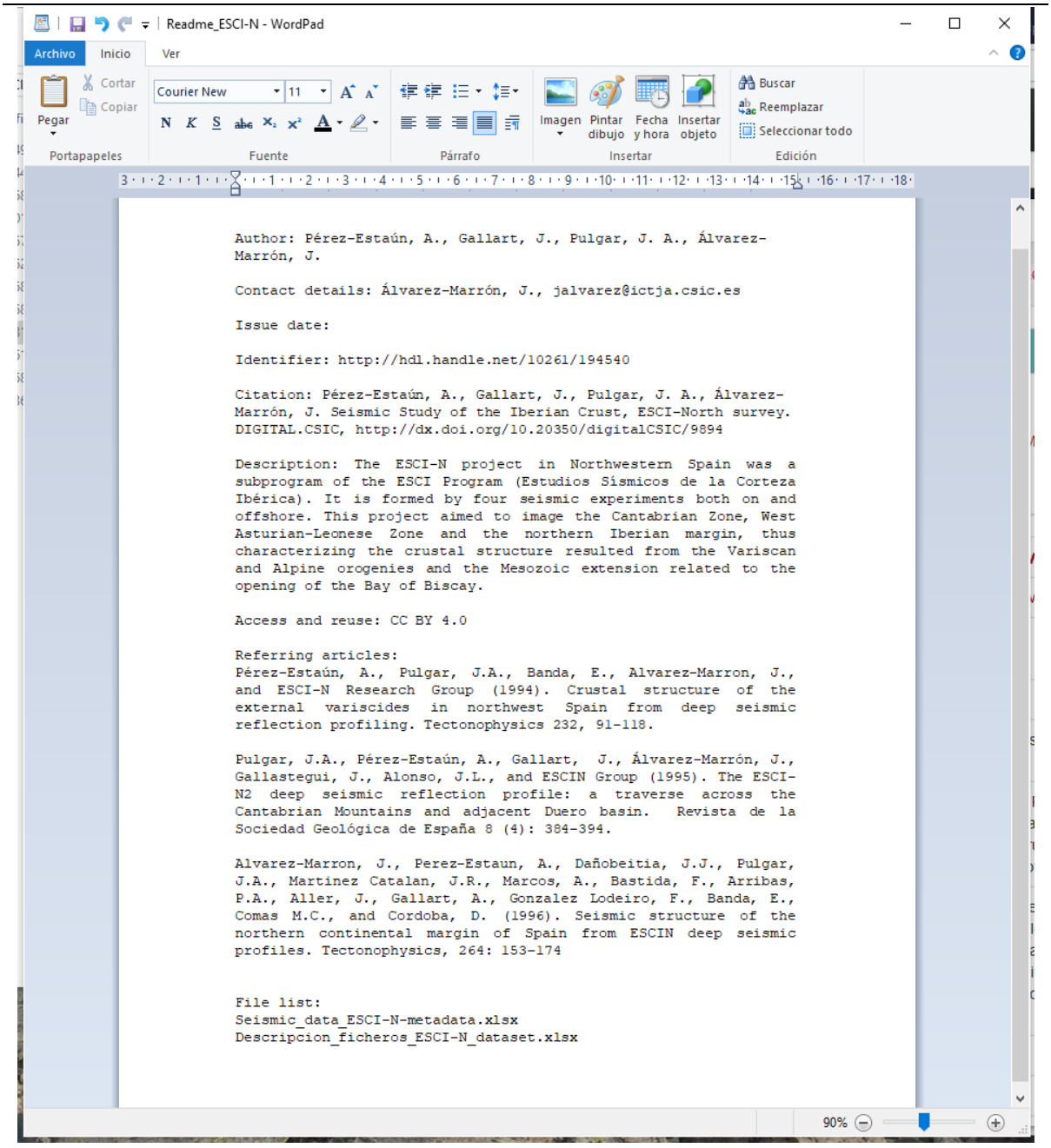
Our database consists of active seismic reflection experiments carried out since the 90's in the Iberian Peninsula. The data were acquired onshore and offshore and at different scales, from continental/crustal scale sampling the Variscan and Alpine orogens, to local/exploration scale in different geographical and geological settings.

According to the statistics, our dataset is being used more and more by users all around the world. Interestingly, our database is being visited mainly by users from USA and Canada, followed by European users and being China the third country in number of downloads of any of all the projects included in the ICTJA-CSIC dataset.

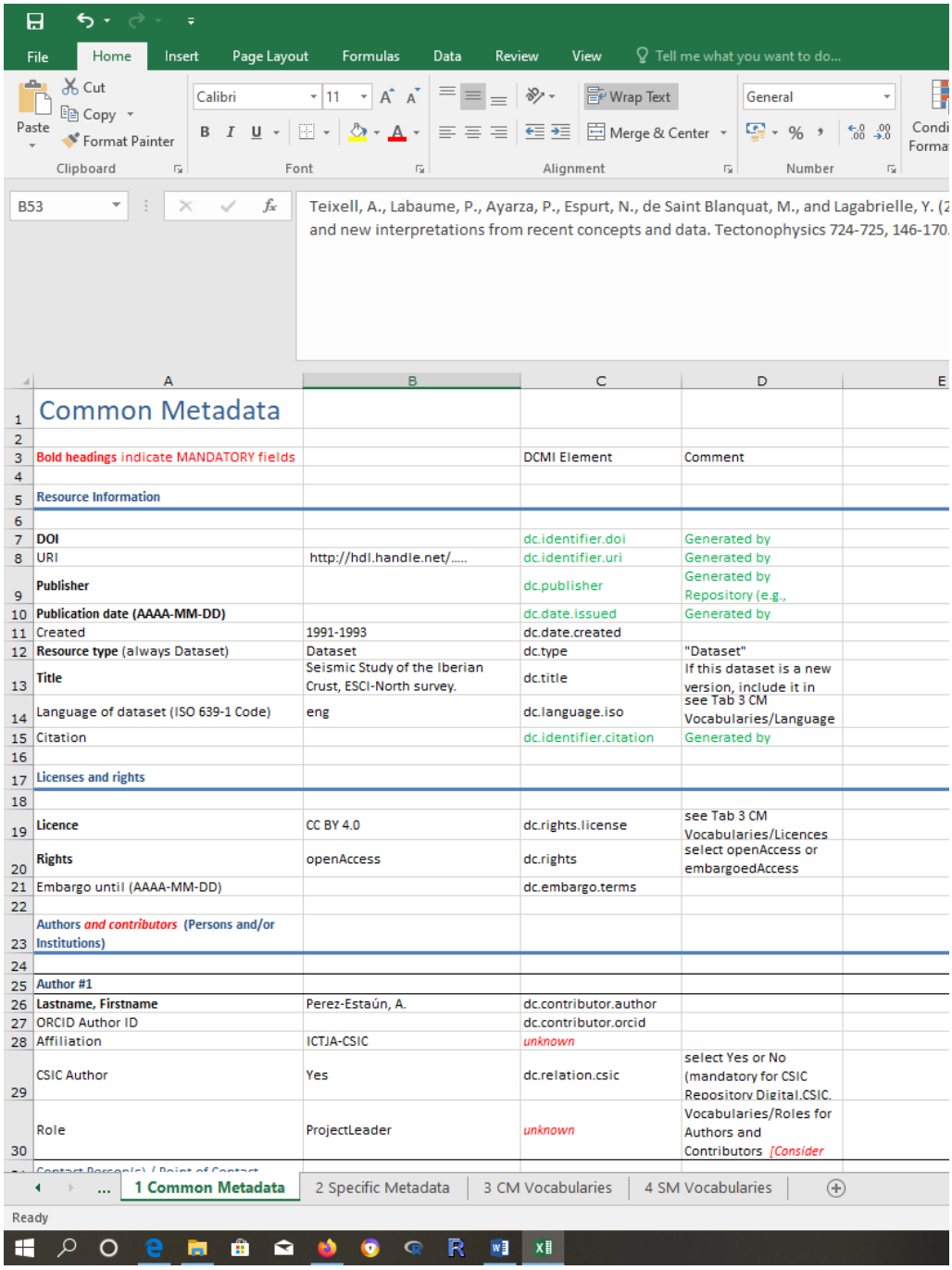
Although this dataset is achieving the national and international mandates for open data, there is still a lot of work needed to promote open access databases through the Solid Earth Sciences community and specifically, the seismic community. Thus, we are currently working on promoting new international collaborations, disseminate the results of this project and constantly updating our database.

5 Appendices

Annex I: README.txt example of the ESCI-N dataset.



Annex II: common and specific metadata example of the ESCI-N dataset.

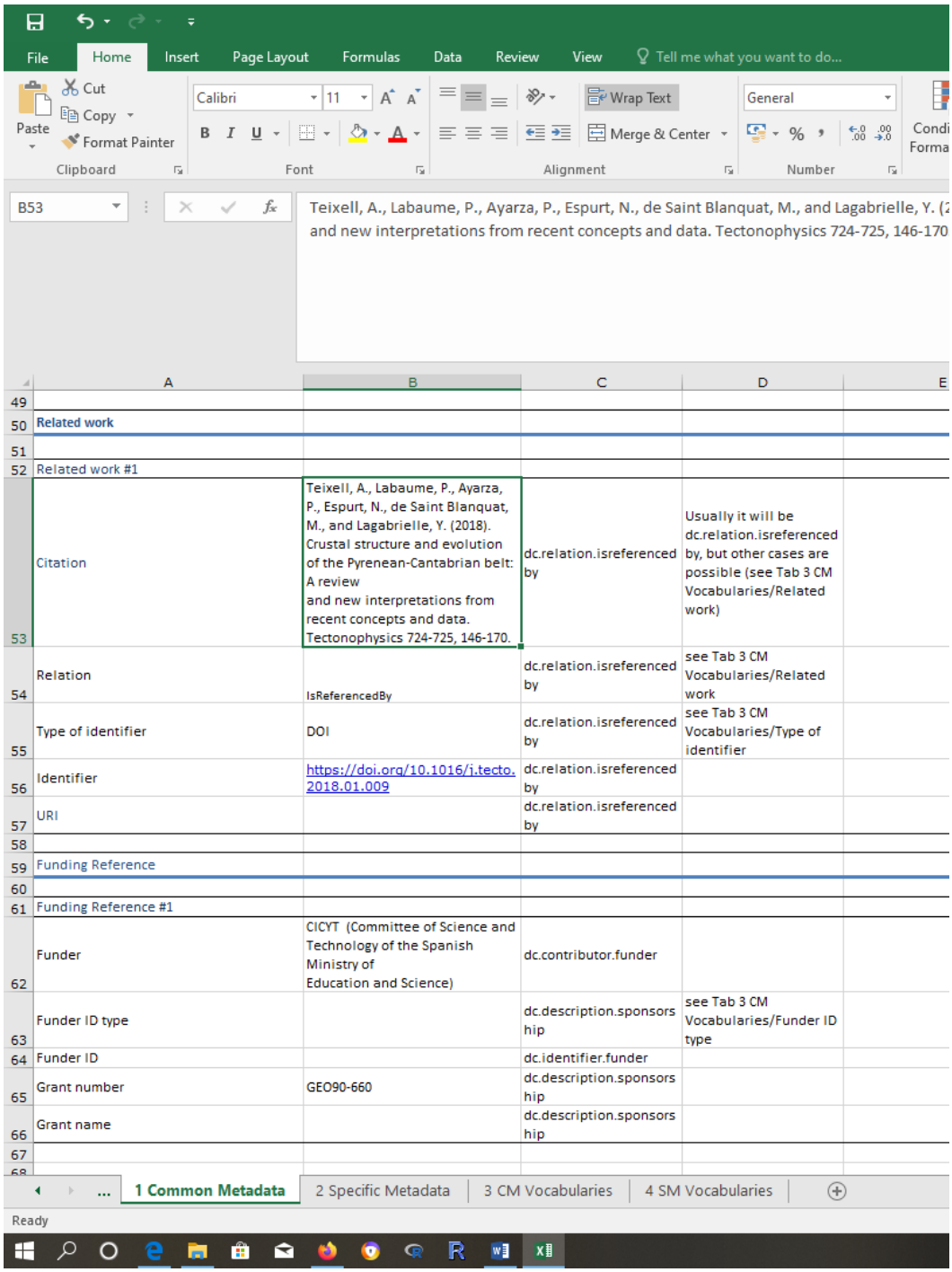


Teixell, A., Labaume, P., Ayarza, P., Espurt, N., de Saint Blanquat, M., and Lagabrielle, Y. (2
and new interpretations from recent concepts and data. Tectonophysics 724-725, 146-170.

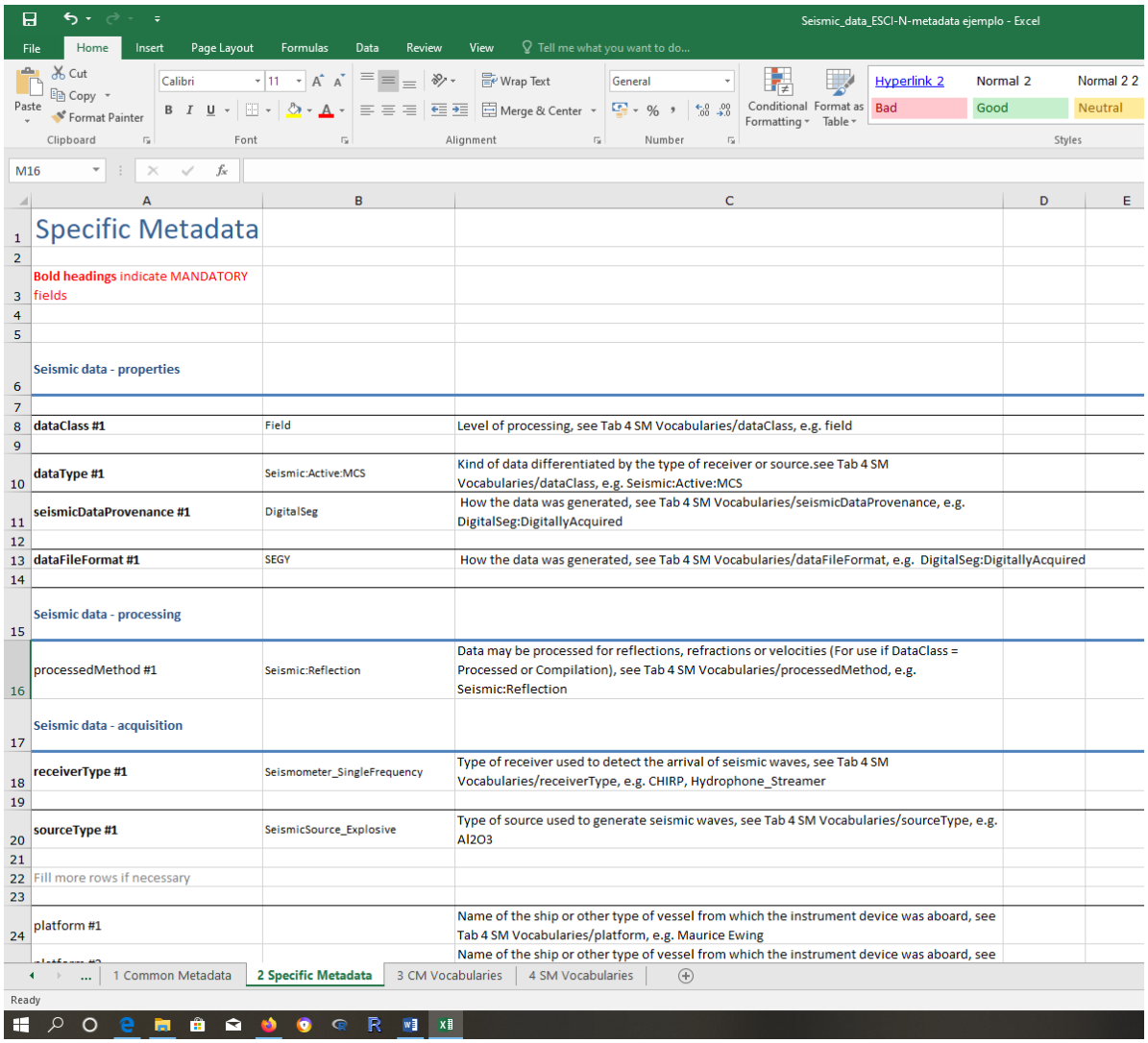
	A	B	C	D	E
1	Common Metadata				
2					
3	Bold headings indicate MANDATORY fields		DCMI Element	Comment	
4					
5	Resource Information				
6					
7	DOI		dc.identifier.doi	Generated by	
8	URI	http://hdl.handle.net/.....	dc.identifier.uri	Generated by	
9	Publisher		dc.publisher	Generated by Repository (e.g.,	
10	Publication date (AAAA-MM-DD)		dc.date.issued	Generated by	
11	Created	1991-1993	dc.date.created		
12	Resource type (always Dataset)	Dataset	dc.type	"Dataset"	
13	Title	Seismic Study of the Iberian Crust, ESCI-North survey.	dc.title	If this dataset is a new version, include it in	
14	Language of dataset (ISO 639-1 Code)	eng	dc.language.iso	see Tab 3 CM Vocabularies/Language	
15	Citation		dc.identifier.citation	Generated by	
16					
17	Licenses and rights				
18					
19	Licence	CC BY 4.0	dc.rights.license	see Tab 3 CM Vocabularies/Licences	
20	Rights	openAccess	dc.rights	select openAccess or embargoedAccess	
21	Embargo until (AAAA-MM-DD)		dc.embargo.terms		
22					
23	Authors *and contributors* (Persons and/or Institutions)				
24					
25	Author #1				
26	Lastname, Firstname	Perez-Estaún, A.	dc.contributor.author		
27	ORCID Author ID		dc.contributor.orcid		
28	Affiliation	ICTJA-CSIC	*unknown*		
29	CSIC Author	Yes	dc.relation.csic	select Yes or No (mandatory for CSIC Repository Digital.CSIC.	
30	Role	ProjectLeader	*unknown*	Vocabularies/Roles for Authors and Contributors *[Consider*	

1 Common Metadata | 2 Specific Metadata | 3 CM Vocabularies | 4 SM Vocabularies

Ready



	A	B	C	D
49				
50	**Related work**			
51				
52	Related work #1			
53	Citation	Teixell, A., Labaume, P., Ayarza, P., Espurt, N., de Saint Blanquat, M., and Lagabrielle, Y. (2018). Crustal structure and evolution of the Pyrenean-Cantabrian belt: A review and new interpretations from recent concepts and data. Tectonophysics 724-725, 146-170.	dc.relation.isreferenced by	Usually it will be dc.relation.isreferenced by, but other cases are possible (see Tab 3 CM Vocabularies/Related work)
54	Relation	IsReferencedBy	dc.relation.isreferenced by	see Tab 3 CM Vocabularies/Related work
55	Type of identifier	DOI	dc.relation.isreferenced by	see Tab 3 CM Vocabularies/Type of identifier
56	Identifier	https://doi.org/10.1016/j.tecto.2018.01.009	dc.relation.isreferenced by	
57	URI		dc.relation.isreferenced by	
58				
59	Funding Reference			
60				
61	Funding Reference #1			
62	Funder	CICYT (Committee of Science and Technology of the Spanish Ministry of Education and Science)	dc.contributor.funder	
63	Funder ID type		dc.description.sponsorship	see Tab 3 CM Vocabularies/Funder ID type
64	Funder ID		dc.identifier.funder	
65	Grant number	GEO90-660	dc.description.sponsorship	
66	Grant name		dc.description.sponsorship	
67				

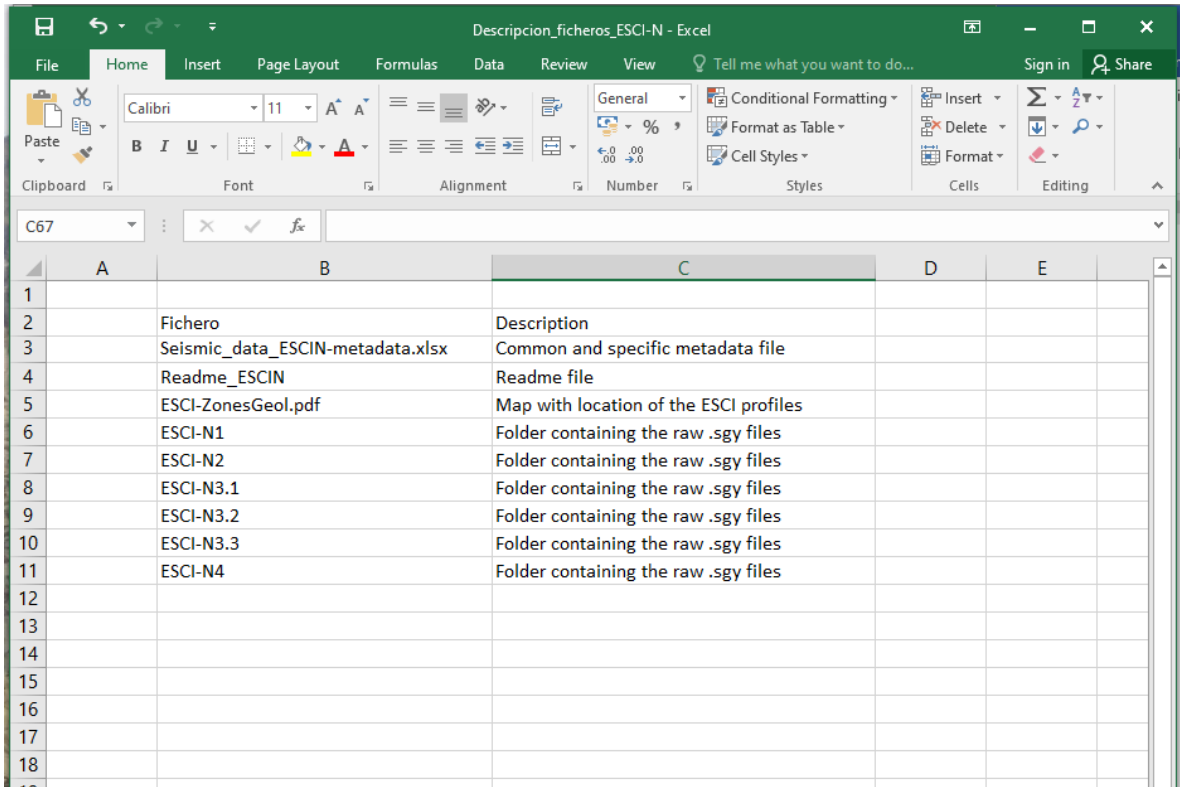


Seismic_data_ESCI-N-metadata ejemplo - Excel

	A	B	C	D	E
1	Specific Metadata				
2					
3	**Bold headings** indicate MANDATORY fields				
4					
5					
6	Seismic data - properties				
7					
8	**dataClass #1**	Field	Level of processing, see Tab 4 SM Vocabularies/dataClass, e.g. field		
9					
10	**dataType #1**	Seismic:Active:MCS	Kind of data differentiated by the type of receiver or source.see Tab 4 SM Vocabularies/dataClass, e.g. Seismic:Active:MCS		
11	**seismicDataProvenance #1**	DigitalSeg	How the data was generated, see Tab 4 SM Vocabularies/seismicDataProvenance, e.g. DigitalSeg:DigitallyAcquired		
12					
13	**dataFileFormat #1**	SEGY	How the data was generated, see Tab 4 SM Vocabularies/dataFileFormat, e.g. DigitalSeg:DigitallyAcquired		
14					
15	Seismic data - processing				
16	processedMethod #1	Seismic:Reflection	Data may be processed for reflections, refractions or velocities (For use if DataClass = Processed or Compilation), see Tab 4 SM Vocabularies/processedMethod, e.g. Seismic:Reflection		
17	Seismic data - acquisition				
18	**receiverType #1**	Seismometer_SingleFrequency	Type of receiver used to detect the arrival of seismic waves, see Tab 4 SM Vocabularies/receiverType, e.g. CHIRP, Hydrophone_Streamer		
19					
20	**sourceType #1**	SeismicSource_Explosive	Type of source used to generate seismic waves, see Tab 4 SM Vocabularies/sourceType, e.g. Al2O3		
21					
22	Fill more rows if necessary				
23					
24	platform #1		Name of the ship or other type of vessel from which the instrument device was aboard, see Tab 4 SM Vocabularies/platform, e.g. Maurice Ewing		
			Name of the ship or other type of vessel from which the instrument device was aboard, see		

1 Common Metadata | 2 Specific Metadata | 3 CM Vocabularies | 4 SM Vocabularies

Annex III: description of the metadata for the ESCI-N dataset.



	A	B	C	D	E
1					
2		Fichero	Description		
3		Seismic_data_ESCIN-metadata.xlsx	Common and specific metadata file		
4		Readme_ESCIN	Readme file		
5		ESCI-ZonesGeol.pdf	Map with location of the ESCI profiles		
6		ESCI-N1	Folder containing the raw .sgy files		
7		ESCI-N2	Folder containing the raw .sgy files		
8		ESCI-N3.1	Folder containing the raw .sgy files		
9		ESCI-N3.2	Folder containing the raw .sgy files		
10		ESCI-N3.3	Folder containing the raw .sgy files		
11		ESCI-N4	Folder containing the raw .sgy files		
12					
13					
14					
15					
16					
17					
18					

6 References

Bernal, I. (2011). Digital.CSIC: Making the Case for Open Access at CSIC, Serials Review, 37:1, 3-8. https://doi.org/10.1080/00987913.2011.10765338.

Bernal, I. and Oficina Técnica de Digital.CSIC (2019). Cómo usar DIGITAL.CSIC para aumentar la presencia de la investigación del ICM en la web. http://hdl.handle.net/10261/180683

Carbonell, R., DeFelipe, I., Alcalde, J., Ivandic, M., and Roberts, R. (2020). Towards an Open Access European Database for Deep Seismic Sounding data. EGU2020-7929. https://doi.org/10.5194/egusphere-egu2020-7929.

DeFelipe, I., Alcalde, J., Carbonell, R., Ivandic, M. and Roberts, R. (2019a). Towards an Open Access Data Policy for Deep Seismic Sounding data. SEISMIX2020

DeFelipe, I., Alcalde, J., Fernandez-Turiel, J. L., Diaz, J., Geyer, A., Molina, C., Bernal, I., Fernandez, J., Carbonell, R. (2019b). Multi-disciplinary data contribution to EPOS e-infrastructure. SEISMIX 2020

DeFelipe, I., Alcalde, J., Fernandez-Turiel, J. L., Diaz, J., Geyer, A., Molina, C., Bernal, I., Fernandez, J., Carbonell, R. (2020). The Spanish node of the multidisciplinary integrated e-infrastructure EPOS. EGU2020-7692. https://doi.org/10.5194/egusphere-egu2020-7692

DeFelipe, I., Atakan, K., Ivandic, M., and Carbonell, R. (2019c). WP 5.4: networking deep seismic data and products. SERA project. SERA Annual Science Meeting, Edinburgh. http://hdl.handle.net/10261/183321.

Mooney, W. D., Prodehl, C., and Pavlenkova, N. I. (2002). Seismic velocity structure of the continental lithosphere from controlled source data. International handbook of Earthquake and Engineering seismology, vol. 81A, ISBN: 0-12-440652-1.

Contact

Project lead	ETH Zürich
Project coordinator	Prof. Dr. Domenico Giardini
Project manager	Dr. Kauzar Saleh
Project office	ETH Department of Earth Sciences

Sonneggstrasse 5, NO H-floor, CH-8092 Zürich
sera_office@erdw.ethz.ch
+41 44 632 9690
Project website www.sera-eu.org