Data Management Plan

University of Urbino

Luis Carlos Castillo-Tellez

2022-10-30

0 Administrative data

Version of DMP: v1.0

Project title: *Bibliometric Analysis of European Research on Digital Divide: An Exploration of the Corporate Landscape*

Start and end of project: 01.11.2022 - 15.02.2023

Data summary: his bibliometric aims to examine the current state of the art of European research in the field of the digital divide by combining the Web of Science, Scopus, and Dimensions bibliographic platforms. Additionally, this work seeks to explore the corporate digital divide. The limit of the search comprises authors with European affiliations within the business, management, economics, technology, and computer science disciplines. A total of 1914 documents, including articles, book chapters, conferences, and proceedings, were found in the three bibliographical sources. The results will be obtained by operating the R programming language using the bibliometrix package and the biblioshiny application.

1 Data Description

How does your project generate new data? Is existing data reused? Which data types (in terms of data formats like image data, text data or measurement data) arise in your project and in what way are they further processed? To what extent do these arise or what is the anticipated data volume?

The source of the data are the Web of Science, Scopus and Dimensions platforms The data will be collected by conducting a specific search in the Web of Science and Scopus. For Dimensions the data will be requested through an API in this case a query will be written according /... This will result in texts (e.g. questionnaires, transcripts)/ numeric data/ video/ audio.... The formats generated will be readable using word processing programm/ statistical software/ video players/ audio player... .

Web of Science

- **Searched Fields:** Keywords and Title
- **Searched Text:** "digital divide*" OR "digital inequalit*" OR "digital gap*"
- **Document Types:** Articles, proceedings book chapters, review articles and early access
- **Web of Science Categories:** Computer science and technology, management, business and economics.
- **Region:** Countries in Europe
- **Time Frame:** 2000 - 2022
- **Total Documents:** 1032
- **Query Link:** Go to Web of Science Query Link
- **Downloaded Data Type:** .txt
- **Download data and Size:** eu_wos_1_500.txt 2.5 MB and eu_wos_501_1032.txt 2.8 MB

Scopus

- **Searched Fields:** Keywords and Title
- **Searched Text:** "digital divide*" OR "digital inequalit*" OR "digital gap*"
- **Document Types:** Articles, proceedings, book chapters, review articles, and early access articles.
- **Scopus Categories:** Computer science and technology, management, business and economics.
- **Region:** Countries in Europe.
- **Time Frame:** 2000 - 2021
- **Total Documents:** 1786
- **Query Link:** Vizualize query
- **Downloaded Data Type:** .csv
- **Download data and Size:** eu_scopus_1_1786.csv 16.9 MB

Dimensions

- **Searched Fields:** Title and Abstract.
- **Searched Text:** "digital divide*" OR "digital inequalit*" OR "digital gap*"
- **Document Types:** Articles, proceedings, and book chapters.
- **Diemensions Categories:** (38) Economics, (35) Commerce, Management, Tourism and Services, and (46) Information and Computing Sciences.
- **Region:** Countries in Europe.
- **Time Frame:** 1999 - 2021
- **Total Documents:** 3108
- **Query Link:** Vizualize query
- **Downloaded Data Type:** .csv
- **Download data and Size:** dim_1_1467.csv 8.3 MB and dim_1468_3108.csv 3.8 MB

During the data processing stage the program xy/yz/zx will be used and the formats of the processed data will be .xxx, .yyy, .zzz.

The volume of the generated data will be estimated at x GB/ TB.

Furthermore, these research data will be reused: Presse- und Informationsamt der Bundesregierung, Berlin (2021). Trust in State and Society during the Corona Crisis (March 2021). GESIS Data Archive, Cologne. ZA7698 Data file Version 1.0.0,

https://doi.org/10.4232/1.13803.

2 Documentation and Data Quality

What approaches are being taken to describe the data in a comprehensible manner (such as the use of available metadata, documentation standards or ontologies)? What measures are being adopted to ensure high data quality? Are quality controls in place and if so, how do they operate? Which digital methods and tools (e.g. software) are required to use the data?

The metadata standard Data Documentation Initiative will be used, as this is the standard followed by the repository [Gesis/ Qualiservice/ etc.] which will store the project' s data. There will be a DDI codebook to describe the study, the data files and the variables. Keywords will be chosen from the European Language Social Science Thesaurus to describe the data. To ensure that the output of the data collection process will result in high-quality, valid data that can be replicated and reused, the following measures will be taken: • Pretests • Data entry validation • Peer review of data • Repeat measurements • Intercoder reliability measures • ... MAXQDA/ Word/ R-Studio/ SPSS/ Python Version xy will be used for processing and analysing of the data. Later, the publication of the data will be (additionally) in .xy formats to ensure long-term usability and reusability for other scientists, so any word processing program/ statistical program/ etc.

can be used for the reuse of the data. The project team follow their data naming convention and folder structure to manage the data responsibly. The versioning of the data will be documented in the processing code/ Git will be used for versioning of the research data.

3 Storage and technical archiving

How is the data to be stored and archived throughout the project duration? What is in place to secure sensitive data throughout the project duration (access and usage rights)?

Nonidentifiable data Throughout the project duration the nonidentifiable data will be stored on the collaborative workspace (R-Studio/ GitHub/ Nextcloud/ Seafile/ etc.). The data will be backed up (daily/ weekly/ fortnightly) on the central backup server of the University.

Identifiable/ sensitive data There will be identifiable/ sensitive data which will be stored encrypted on the institutional system Seafile/ on the Netzlaufwerk of the University of Bremen. The encrypted identifiable data will be backed up (daily/ weekly/ fortnightly) on the central backup server of the university. Only person x and person y/ Only directly involved researchers will have access to identifiable/ sensitive data.

The transcripts will be pseudonymized/ anonymized at the earliest stage possible. The survey data will be anonymized at the earliest stage possible. Survey data/ transcripts and identifiable data (e.g. email address, phone numbers) will be stored separately.

4 Legal obligations and conditions

What are the legal specifics associated with the handling of research data in your project? Do you anticipate any implications or restrictions regarding subsequent publication or accessibility? What is in place to consider aspects of use and copyright law as well as ownership issues? Are there any significant research codes or professional standards to be taken into account?

Data collected in this research project is owned by ... represented by ... .

As there will be personal data, there will be some implications for the publication and accessibility:

Any identifiable personal data will be pseudonymized/ anonymized before allowing others to reuse the data.

There will be an informed consent form for the participants regarding to the preservation in a repository and the scientific reuse of the pseudonymized data after the end of the project.

Pseudonymized data will be made accessible only for the scientific use.

vulnerable person/ groups

The entire dataset will not made public to ensure the confidentiality and safety of the participants. As the participants are a vulnerable group/ As the risk of re-indification is high/ as the expected harm in case of a re-indification is high, the data will be stored under protected access in the e. g. Qualiservice for 10 years and made accessible only for the scientific use.

5 Data exchange and long-term data accessibility

Which data sets are especially suitable for use in other contexts? Which criteria are used to select research data to make it available for subsequent use by others? Are you planning to archive your data in a suitable infrastructure? If so, how and where? Are there any retention periods? When is the research data available for use by third parties?

The “milestones” of the research data/ The research data in their final versions (e.g. anonymized survey data/ pseudonymized transcripts, code, informed consent form, questionnaire, interview guide) will be stored in a

disciplin-specific repository (e.g. Gesis/ Qualiservice) for the long-term preservation and for the (scientific) reuse.

The anonymized data will be published with the license xy at the end of the project in the repository xy.

The pseudonymized data will be accessible for the scientific reuse with the license xy ("conditional access"/ protected access"). The access conditions will be described in the repository xy.

The data (e.g. transcripts) will not be accessible for the reuse because of the vulnerable groups of participants. Interview guide and codebook will be published with license xy in the repository xy.

The encrypted video files/ the sensitive data will be stored on an institutional server for 10 years under "protected access". Metadata and conditions for access to and reuse of the data will be published in the repository xy.

Following material will be published to make the accessible/ open data understandable and reusable for other scientists: Methods report, instrument of data collection (e.g. questionnaire, code), codebook, code for data processing and data analysis, informed consent form, tools and software including version number.

The papers will include information and accessibility of the data to make them findable. Furthermore, keywords in the research data repository will make the data more findable.

The research data will be published the latest at the end of the project/ before the first publication.

There will be a rentention period of x years, because of potential patent rights.

6 Responsibilities and resources

Who is responsible for adequate handling of the research data (description of roles and responsibilities within the project)? Which resources (costs; time or other) are required to implement adequate handling of research data within the project? Who is responsible for curating the data once the project has ended?

Project coordinator: _____
Name, affiliation, email address, ID's (e.g. ORCID)

Principal investigator: _____
Name, affiliation, email address, ID's (e.g. ORCID)

Author of DMP: _____
Name, affiliation, email address, ID's (e.g. ORCID)

Data officer and responsible for DMP:_____ Name, affiliation, email address, ID's (e.g. ORCID)

The estimated costs of RDM-activities:

Writing DMP: [2 hours - 2 days] Transcription: 1 hour interview = 4,5 hours transcription per interview; x interviews X hour wage of research assistant = xxxx Euro Codebook: x hours/ days Methods report: x hours/ days

Costs for archiving in repository xy: xxx Euro (offer attached to proposal)

Once the project has ended the data will be curated in the repository xy.

Once the project has ended the data on the institutional server will be curated/ made accessible by [. . .].