Data Management Plan

University of Urbino

Luis Carlos Castillo-Tellez

2022-11-09

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: This bibliometric aims to examine the 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. After processing, merging, and cleaning,
a total of 1914 documents, unique documents were incorporated into the final data set, 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

This research project will use secondary data collected by conducting a specific search on the digital divide using the Web of Science, Scopus, and Dimensions platforms. The three platforms have different graphic user interfaces that delimit the search and the supported formats in which bibliographical data is downloaded. The content of bibliographical data varies between text, numeric, and integers data types, and the formats that generate each platform will be readable using the R programming language

1.1 Data collection

During this stage, the search criteria, the queries, and the formats downloaded are the following:

Web of Science This platform allows one to choose the formats and the fields. While exporting the data, a *plain text file* was chosen. The custom selection used all the Web of Science core collection fields for the record content.

- **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:** 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

This database also allows one to choose the formats and the necessary fields to conduct a bibliometric analysis. In this case, all the categories of citation information, bibliographical information, abstracts, and keywords.

- **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:** Vizualize query
- **Downloaded Data Type:** .csv
- **Download data and Size:** eu_scopus_1_1786.csv 16.9 MB

Dimensions

This database has a less developed graphic user interface. Even though it lets one choose the format, it does not let one choose the fields. However, the search can be customized using the Application Programming Interface API.

- **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:** Not available for graphic user interface.
- **Downloaded Data Type:** .csv
- **Download data and Size:** dim_1_1467.csv 8.3 MB and dim_1468_3108.csv 3.8 MB

1.2 Data Processing

The R programming language environment will be used during the data processing stage. After downloading, the raw files (.txt and .csv formats) from each platform will be converted into a bibliographic database format using the bibliometrix package. The volume of the generated data in this repository will be estimated between 0.7 to 1 GB.

After converting the downloaded datasets from the three platforms into bibliographic data frames, differences were found in column length and names. For example, the converted data frame from the Web of Science contained 73 variables, Scopus 37, and Dimensions 30. To solve this problem, first, we followed the Web of Science Core Collection field tags to homogenize the variable's names. Second, the bibliometrix manual suggests selecting the main variables to conduct a bibliometric analysis. As a result, the final bibliographic data frame contains 29 variables.

This script contains the data processing, cleaning, and merging from the three sources.

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. 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.

4 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)