Data Management Plan
University of Urbino

Luis Carlos Castillo-Tellez

2022-11-10

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 data from 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 unique documents were incorporated into the final data set, including articles, book chapters, conferences, and proceedings. 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

Before collecting the data, access to the graphical user interfaces of each platform was granted through university affiliation. 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

- **Date of Search:** 10.10.2022
- **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
- **Date of Search:** 05.10.2022
- **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.
- **Dimensions Categories:** (38) Economics, (35) Commerce, Management, Tourism and Services, and (46) Information and Computing Sciences.
- **Region:** Countries in Europe.
- **Time Frame:** 1999 - 2021
- **Date of Search:** 14.10.2022

- **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, there were differences in column length and names, meaning the three platforms used different codebooks for naming the variables. For example, the Web of Science codebook contained 73 variables, the Scopus codebook 37 variables, and the Dimensions codebook 30 variables. To solve this problem, first, we followed the Web of Science Core Collection codebook 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 (see codebook).

This script contains the R-Studio workflow used to process, clean, and merge the bibliographic data frames from the three sources.

2. Documentation and Data Quality

The metadata creation will follow the Data Documentation Initiative DDI standards, including a codebook that describes the data and the variable types. To guarantee replicability both in the data collection and output, the reuse of data, and high-quality results, the following measures will be taken: pretests, data validation, peer review of data, repeat measurements, and the use of software and digital tools. This repository, bibliometric_digital_divide_Europe, will follow the R-Studio workflow using the bibliometrix package and the biblioshiny application for data processing and analysis. An ordered structure of the folders with documented code will also be managed using the version control offered by GitHub to track changes and manipulate data responsibly.

3. Storage and technical archiving

This research project does not contain any sensible data; thus, throughout the project duration, data will be stored on the collaborative workspace R-Studio, GitHub, and Dropbox.

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.

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)