
‘DIGITAL SOVEREIGNTY BY DESIGN’

DIGITAL SOVEREIGNTY DATASET

CODEBOOK

ACADEMIC ARTICLES AND POLICY DOCUMENTS DATA

NEWCASTLE UNIVERSITY
NORTHUMBRIA UNIVERSITY. NEWCASTLE
2023

Table of Contents

1	Introduction & Citation	2
1.1	Project information	2
1.2	Project team	2
1.3	Citation	2
1.4	Note on the terms of use of these datasets	3
2	Data collection protocol	4
2.1	Academic Articles	4
2.1.1	Criteria	4
2.2	Policy Documents	4
2.2.1	Criteria	5
2.3	Important note on limitations of the data collection	5
3	Overview	6
3.1	Description of academic articles dataset	6
3.2	Description of policy documents datasets	6

About This Codebook

This Codebook presents text data stemming from academic articles (Scopus and WoS) and Policy documents collected from EU institutional repositories (European Commission, European Parliament, Council of the European Union). The data was collected from January to May 2023, as part of the "Digital Sovereignty by Design" research project framework. The primary purpose of this data collection is to conduct topic modelling in order to identify current debates on digital sovereignty through international relations and political science lenses.

1 Introduction & Citation

1.1 Project information

The 'Digital Sovereignty by Design' project aims to 1) engage with current law and policy approaches to cybersecurity in the United Kingdom and in the European Union, 2) identify how the EU's current shift towards a digital sovereignty approach may negatively impact upon UK private-sector technology providers, in particular microprocessors, and 3) consider ways in which these negative impacts may be mitigated.

In order to map the current state of the art in digital sovereignty, we carried out not only an analysis of the academic literature, but also collected text data from academic articles and from EU policy documents using quantitative analysis of texts (namely topic modelling - probabilistic (generative) modelling - Latent Dirichlet Allocation - LDA). The text data is contained in these datasets.

Additional information about the project is available from the website:

<https://www.digitalsovereigntyby.design>

1.2 Project team

The project's research team consists of Newcastle University and Northumbria University (Newcastle Campus) staff.

The project's Principal Investigator (PI) is Professor in Law & Emerging Technologies at Newcastle University (Law School) - [Ben Farrand](#).

Co-PI is Professor in International Relations and European Politics at Northumbria University (Department of Social Sciences, ADSS) - [Helena Farrand Carrapico](#).

The project's Research Assistant is a Postdoctoral Fellow in Cybersecurity and Technology Policy at Northumbria University (Department of Social Sciences, ADSS) - Dr [Aleksei Turobov](#).

1.3 Citation

Scholars who wish to use these datasets in their research are kindly requested to use the following citation:

Turobov, Aleksei and Farrand, Benjamin and Farrand Carrapico, Helena (2023).
*"Digital Sovereignty by Design", Academic Articles and Policy Papers, Digital
Sovereignty Dataset*. Colchester, Essex: UK Data Service.

1.4 Note on the terms of use of these datasets

These datasets are open and available, free of charge and do not require you to register your data. You can use them for your analysis, graphs, teaching, and other academic-related and non-commercial purposes. We ask our users to always cite the data and our datasets correctly. We do not allow other uses of these data, including but not limited to redistribution, commercialisation and other for-profit usages. If a user is interested in such a use, or has doubts about the license, they will have to refer to the original source and check with them if this is allowed and what requirements they need to fulfil.

2 Data collection protocol

We collected data from two distinct types of sources: (1) Text data stemming from academic articles (digital sovereignty); (2) Text data stemming from policy documents produced by three of the main EU institutions.

The main purpose of this data collection is to cover all discussions regarding the digital sovereignty concept within the academic and policy fields.

2.1 Academic Articles

One dataset was created concerned with the concept of digital sovereignty. The collected articles mainly stem from International Relations, European Union Studies, Science and Technology Studies, and Law.

2.1.1 Criteria

The collection was carried out taking into account the following criteria:

- we collected only peer-reviewed articles;
- the collection was limited to English language articles;
- the time period was not set and was determined by the available articles: for Web of Science (WoS) from 2014 to 2023 and Scopus from 2013 to 2023;
- the search was based on the keyword "digital sovereignty"

It is worth noting that the resulting time frame indicates that the phenomenon of "digital sovereignty" is relatively new and is just beginning to be actively studied. As a result, 101 articles from the WoS database and 112 articles from Scopus were collected. Some articles were present in both databases, and were manually verified and deleted to avoid duplication. Thus, the complete database of collected articles comprises 156 articles and includes bibliographic information about the article, annotation/abstract and the keywords from the article itself.

2.2 Policy Documents

The collection of policy documents texts-data was carried out from open databases (repositories) of the three main EU institutions: the European Commission, the European Parliament, and the Council of the European Union. For uniformity of data collection, uniform selection criteria were applied.

2.2.1 Criteria

- keywords: "digital sovereignty", "data sovereignty", "technology sovereignty", "cyber sovereignty";
- the collection was limited to English language documents;
- the collection included all types of documents (except tables, Amendments and graphs);
- the time period was not set;
- in the case of large documents and reports, the abstract, the introduction, the executive summary, or the conclusion were collected;

As a result, the European Commission dataset contains 61 texts (from the Register of Commission Documents) and 158 texts (from the Repository of most commonly used public documents) from 2009 to 2023; the European Parliament dataset contains 368 texts from 2011 to 2023; and the Council of the European Union dataset contains 221 texts from 2001 to 2023.

2.3 Important note on limitations of the data collection

The usage of repositories and the functional performance of EU institutions websites deserves special attention. These datasets do not guarantee the presence of all texts for the specified keywords. Some of the texts at the time of collection were not available online, and some were not allowed to be copied (for example, a scanned copy was posted on the websites). At the time of the data collection, the European Commission repository worked intermittently and did not display documents correctly, as well as constantly changing search results (this is why we expanded our dataset by collecting data from the "Repository of most commonly used public documents"). Thus, when working with these datasets, the indicated limitations in the volume of texts-data and their completeness is worth bearing in mind.

3 Overview

This Codebook presents five datasets. One dataset contained academic articles data on digital sovereignty, and four other datasets collected policy documents from EU institutions. The academic articles' datasets contain complete bibliographic information, abstract and keywords (if available). The datasets for the policy documents include the document's title and its complete text.

3.1 Description of academic articles dataset

Table 1: Data Articles

Keyword	Digital Sovereignty
Time frame	2013 - 2023
Criteria	Articles; English; Web of Science; Scopus
Subjects	all
Total	156

3.2 Description of policy documents datasets

Table 2: Data Policy Documents

Institution	European Commission	European Parliament	Council of the European Union
Keywords	Cyber sovereignty, Data sovereignty, Technology sovereignty, Digital sovereignty		
Criteria	All types of documents (except tables, Amendments and graphs) English language		
Time frame	2009 - 2023	2011 - 2023	2001 - 2023
	61 from Register; 158 from Commonly used documents		
Total	219	368	221