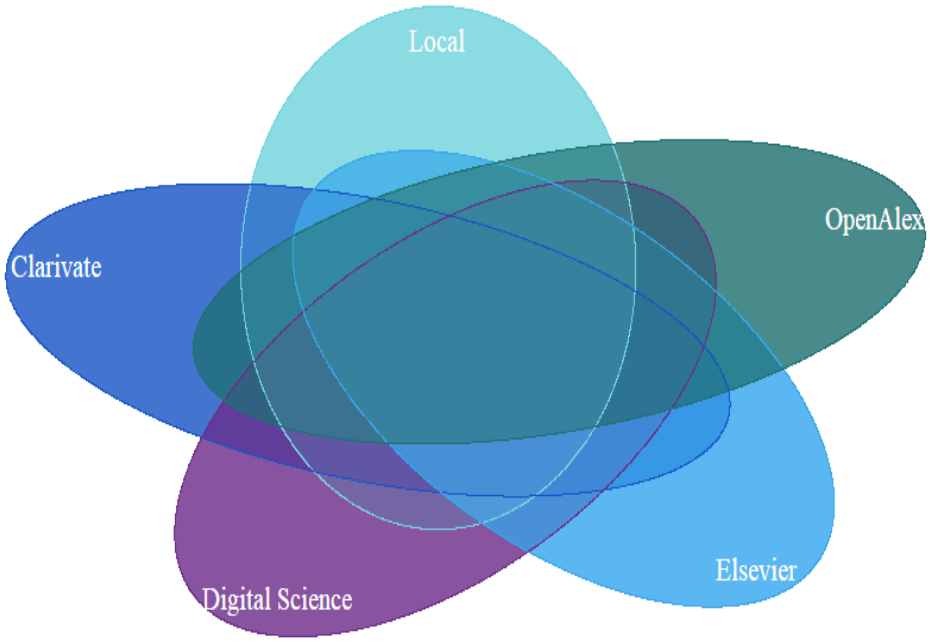
COVERAGE OF DANISH RESEARCH PUBLICATIONS

Local and Global Data Sources of Research Portal Denmark



NORA Team
Final Draft
January 2025

FUNDING ACKNOWLEDGEMENT

Like all other parts of the development and operation of the Research Portal Denmark, this report has been financed by the Danish Agency for Higher Education and Science.

CONTRIBUTORS IN ALPHABETICAL ORDER AND THEIR ROLES

Specified using the **CRediT Taxonomy** contributor roles:

Marija Rakas – 0000-0003-3050-1204 – marak@dtu.dk
ROLES: Data Curation; Investigation; Writing – Original Draft; Visualization
AFFILIATION: National Open Research Analytics, Technical University of Denmark

Nikoline Dohm Lauridsen – 0000-0002-6139-0108 – nidl@dtu.dk
ROLES: Data Curation; Investigation; Writing – Original Draft; Visualization
AFFILIATION: National Open Research Analytics, Technical University of Denmark

Karen Hytteballe Ibanez – 0000-0002-8229-0392 – kshi@dtu.dk
ROLES: Writing – Review & Editing; Supervision
AFFILIATION: National Open Research Analytics, Technical University of Denmark

Mogens Sandfær –0000-0001-8436-5346 – mosa@dtu.dk
ROLES: Writing – Review & Editing; Supervision
AFFILIATION: National Open Research Analytics, Technical University of Denmark

Dicte Madsen – 0000-0003-4746-4235 – dbbdm@dtu.dk
ROLES: Data Curation, Writing – Review & Editing
AFFILIATION: National Open Research Analytics, Technical University of Denmark

Yang Shen – 0009-0009-5522-1111 – yansh@dtu.dk
ROLES: Data Curation; Software
AFFILIATION: National Open Research Analytics, Technical University of Denmark

Brian Lowe – 0000-0002-8143-6345 – brian@ontocale.com
ROLES: Software
AFFILIATION: Ontocale

Franck Falcoz – 0000-0003-3666-5699 – fkybus@gmail.com
ROLES: Data Curation; Software
AFFILIATION: Vox Novitas

Pedro Parraguez Ruiz – 0000-0002-0017-4057 – pedro@dataverz.net
ROLES: Data Curation; Software
AFFILIATION: Dataverz

TABLE OF CONTENTS

1. INTRODUCTION

This report compares publication data from the Danish institutional data sources (Local Data), and four major global data sources: Clarivate, Digital Science, Elsevier, and OurResearch (Global Data). The terms **Local**, and **Global** will be used throughout the report to refer to the two different groups of data sources. Both groups of data sources are data providers of Research Portal Denmark's publication databases.

The first part of the report presents an overall coverage analysis, comparing the data sources by looking at their overlapping and unique publications. Overlapping publications refer to those matched across multiple sources, while unique should be understood as 'unmatched', representing cases where the procedures were not able to confidently create a match across data sources.

The second part of the report delves into the unique publications of each data source, including aspects such as publication types, languages, and subject classification.

The third part of the report presents a coverage analysis comparing each of the Global data sources with the latest data set from the now discontinued Danish National Bibliometric Research Indicator (BFI) from 2021.These three parts together, should support the understanding of research publication coverage across all five databases (Local and Global).

2. METHODOLOGY

The methodology section details the approach used to compile and analyse publication metadata related to Denmark from 2011 to 2023. Section 2.1 describes the acquisition of data from the different data providers. Section 2.2 outlines the matching procedure, followed by the identification and validation of matching records and clusters, while section 2.3 introduces the matching specifics of the BFI part of the analysis. Finally, section 2.4 presents the analytical approach, focusing on the segmentation of publications described by overlapping and unique publications of all five data sources.

2.1 DATA SOURCES AND THE ACQUISITION OF DATA

The coverage report utilizes data from both Danish research institutions (Local Data), and four global data sources (Global Data), to compile a comprehensive set of publication metadata for publications related to Denmark from 2011 to 2023.

Local Data is based on data from 21 Danish research institutions' local systems and portals[1], which are used for registering their own publication outputs and activities (also known as CRIS systems).

Global Data is based on data from the four major global data sources and products:

- Clarivate's Web of Science and InCites databases.
- Digital Science's Dimensions database.
- Elsevier's Scopus and SciVal databases.
- OurResearch's OpenAlex database.

Data selection and extraction of metadata from the sources is explained in the table below. While the scope of the publication databases in Research Portal Denmark is 2011 and onwards, for this report, we

have confined the scope to publications from the years 2011 to 2023 to avoid including the still incomplete coverage of the publication year 2024.

The national standard format DDF-MXD (Danish Research Database Metadata Exchange Format for Documents)[2] format is used when harvesting publication data registered in the local systems and portals of Danish research institutions (Danish Data Providers) for the Research Portal Denmark's database: Local Data[1].

Data in JSON or XML formats from the three data providers, Clarivate, Digital Science, and Elsevier, are imported and made available for the Research Portal Denmark's databases: Global Data - Across All Data[2], Clarivate Data[3], Digital Science Data[4] and Elsevier Data[5]. You can read more about the Research Portal Denmark overall data pipeline and system architecture that collect, enrich and link data from various sources in the full technical documentation[6].

Data in JSON format from OurResearch is imported and currently available only in the Research Portal Denmark's internal prototype database: Open Data.

Table 1: Data selection and extraction from the publication data sources of Research Portal Denmark

Data Source	Data Selection Criteria	Extraction of Metadata
Clarivate	*Web of Science (WoS) Core Collection* (all indexes): Publication Years: 2011-to date Country code: DK	– Bibliographic metadata and data provider's enhancements (e.g., enhanced affiliations, subjects) from WoS API Expanded – Advanced + WoS Starter API received as data dump via email. – Metrics metadata (e.g. OA status) from InCites Document Level Metrics API received as data dump via email. – Data received in XML format and parsed to JSON format.
Digital Science	*Dimensions* Publication Years: 2011-to date Country code: DK	– Bibliographic metadata, data provider's enhancements (e.g., GRID ID to affiliations, subject classifications) and metrics from Dimensions Analytics API. – Data extracted in JSON format.

[1] https://local.forskningsportal.dk
[2] https://global.forskningsportal.dk
[3] https://clarivate.forskningsportal.dk
[4] https://digitalscience.forskningsportal.dk
[5] https://elsevier.forskningsportal.dk
[6] https://forskningsportal.dk/about-data-documentation/technical-documentation/

Elsevier	*Scopus* Publication Year: 2011-to date Country code: DK	– Bibliographic metadata and data provider's enhancements (e.g., affiliation information from OrgDB) received as data dumps via email. – Metrics metadata (e.g., SDG tags) from SciVal. – Data received in XML format and parsed to JSON format.
Local	*Danish Data Providers* Publication Year: 2011-to date Publication types: a selection[7] Country code: DK	– Bibliographic metadata as exposed by the local systems and portals of Danish research institutions (Danish Data Providers) via their OAI-PMH data provider service. – Data harvested in DDF-MXD (Danish Research Database Metadata Exchange Format)[8] format. – Data received in XML format and parsed to JSON format.
OurResearch	*OpenAlex* Publication Year: 2011-to date Country code: DK + List of DOIs with no DK country code in OpenAlex, but identified as DK in Local Data	– Bibliographic metadata and data provider's enhancements extracted via monthly snapshots. – Data received as GZ files, decompressed and decoded to JSON format.

2.2 MATCHING PROCEDURE

The analysis focuses on the Local, and Global data sources as defined in this report, each aiming to provide a comprehensive set of publication metadata: Local Data covers publications pertaining to Danish Data Providers, and the Global Data include publications pertaining to Denmark - all from 2011 to 2023. Consequently, many publications are described by metadata from many data sources.

Identifying these matching metadata records is key to this analysis, as well as for the retrieval and interlinking features of the publication databases of Research Portal Denmark[9].

To identify publications associated with records from more than one Danish Data Provider, a matching procedure has been developed, along with a similar one for the Global Data. Both matching procedures include the following steps:

[7] A selection of publication types and categories include Journal Article, Journal Comment, Journal Review, Newspaper Article, Book, Book Chapter, Book Preface / Encyclopaedia article, Report, Report Chapter, Conference Paper, Conference Abstract, Conference Poster, Working Paper / Preprint, Lecture Notes, Thesis Doctoral, Thesis PhD, and Other.

[8] https://forskningsportal.dk/about-data-documentation/publications/local-data/

[9] This interlinking functionality between the publication databases of Research Portal Denmark is known as 'Matching Records in'.

- Data cleaning and simplification.
- Identification of potential matches.
- Validation of potential matches.

After identifying and validating 'Local' record clusters **within** Local Data and 'Global' record clusters **within** Global Data, the identification of potentially matching publications **between** Local and Global clusters is performed using:

- Persistent identifiers: DOI and PubMed ID.
- Pre-processed titles in combination with ISSN, eISSN, ISBN, eISBN, and publication years.

This process results in the following groups of clusters:

- Matched Local-Global clusters: publications identified in both Local and Global Data.
- Unmatched Local clusters: publications identified only in Local Data.
- Unmatched Global clusters: publications identified only in Global Data.

The matching procedure **between** Local and Global clusters is not necessarily limited to a single 'Local' or 'Global' record cluster and allows matching of multiple records from each source within one cluster. Consequently, in this analysis, some Local publication records are e.g. grouped and counted as one publication rather than multiple. Due to the consolidated nature of the Local publication clusters, more individual Local publication records are used in the Local-Global matching process, allowing for more matches between the Local and Global clusters, than if the consolidated Local clusters were matched with the Global clusters instead.

You can read more about the matching algorithms that connects publication records in Research Portal Denmark in the technical documentation[10].

2.3 BFI MATCHING PROCEDURE

The matching process used for the BFI analysis part of the report relies on the publication IDs from the BFI data set called UUIDs (the original IDs of the publication records from Danish Data Providers' CRIS systems). Initially, these publication IDs were matched with Local Data. Of the 28.764 publications (31.861 records) from the BFI 2021 dataset, 64 publications (80 records) and 52 patents (52 records) could not be matched, leaving a total of **28.648 BFI 2021 publications** for further matching with the four global data sources, as detailed above in section 2.2.

2.4 ANALYTICAL APPROACH

The results of the matching and clustering of records from Local, and Global Data are presented in the following areas (also illustrated by the Venn diagram zones in Figures 1a-1d):

- Publications described by records in **Local** and one of the four **Global** sources.
- Publications described by records **only** in **Local,** or **only** in one of the four **Global** sources.

Please note that data used as basis for the figures and analyses in the report are based on data harvested for Research Portal Denmark in October 2024.

[10] https://docs.google.com/document/d/1JtDzj5f_s_IbbsiFb9CQtjMcXJzwew5-M07Ph_RTgJw/edit#heading=h.mec7crwjuylj

3. COVERAGE ANALYSIS – AN OVERVIEW

3.1 RECORDS RETRIEVED FROM EACH OF THE 5 SOURCES

The total number of records retrieved[11] from each of the five sources is:

Local	698.204 records
Clarivate	389.804 records
Digital Science	392.360 records
Elsevier	390.334 records
Open Alex	485.835 records

In total **2.356.537 records**

When submitted to the matching procedure, these 2.356.537 publication records identified:

In total **955.404 publications** - found in one or more data sources.

Below in section 3.2 an overview of the publication coverage across **Local** and the individual **Global** data sources is provided. **Local** serves as the starting point for this analysis, with pairwise comparisons made to each of the **Global** data sources. **Local** is used as the reference, as it allows for a focused comparison, emphasizing how each Global data source complements or differs from the coverage found in **Local** Data. This section also provides an overview of the number of unique (or unmatched) publications from each source. Chapter 4 further explores the unique publication records from each individual source.

3.2 INTERSECTIONS AND UNIQUE PUBLICATIONS

The matching procedure found 347.683 publications (26.744 per year on average) identified in **Local** and one or more of the **Global** sources, which is **36,4%** of the total number of publications (955.404)[12].

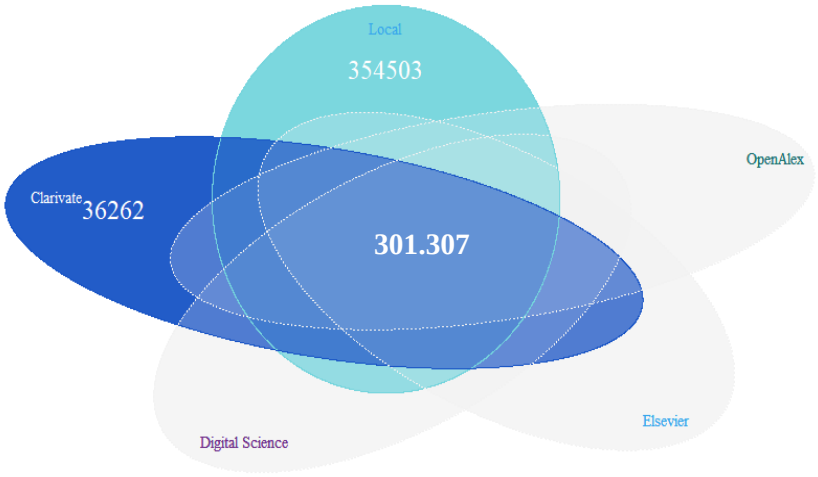
Figure 1a-1d presents a visualization and table 2 below summarizes figures for the publications identified in Local, and each of the four Global sources.

[11] Note that the matching procedure between Local and Global clusters is not limited to a single 'Local' or 'Global' record cluster. Consequently, the total number of records retrieved for the five sources in this report is lower than the results from 2011-2023 publication year queries performed for 'Local Data', 'Clarivate Data', 'Digital Science Data', or 'Elsevier Data'.

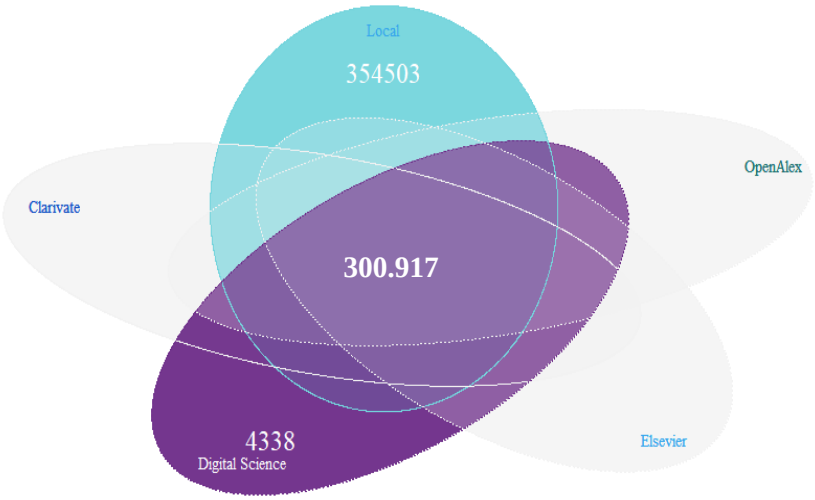
[12] For a full overview of the total number of publications identified by one, two, three or all four sources, please inspect Figure A.1 and Table A.1 in the Appendix section.

Figure 1a-1d: Intersections: Local and the individual Global data sources

The matching procedure identified **301.307** publications (23.177 per year on average) described in both **Local** and **Clarivate**, out of a total of 955.404. This represents **31,5%** of the total number of publications.



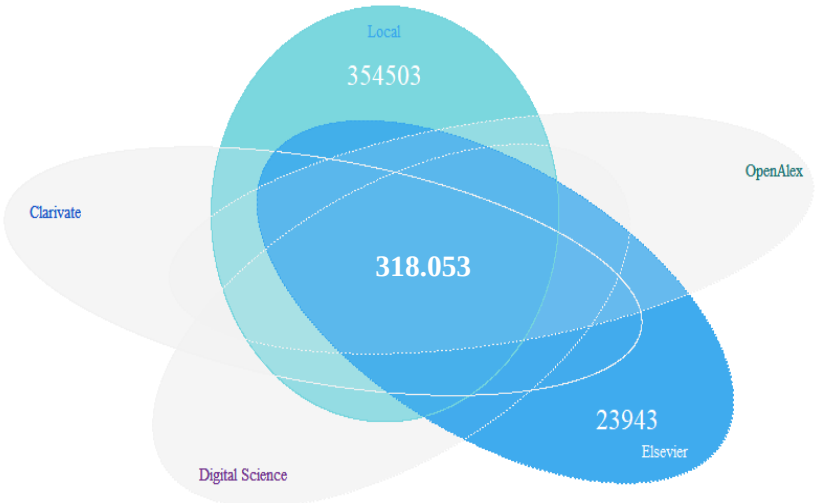
1a: Local versus Clarivate



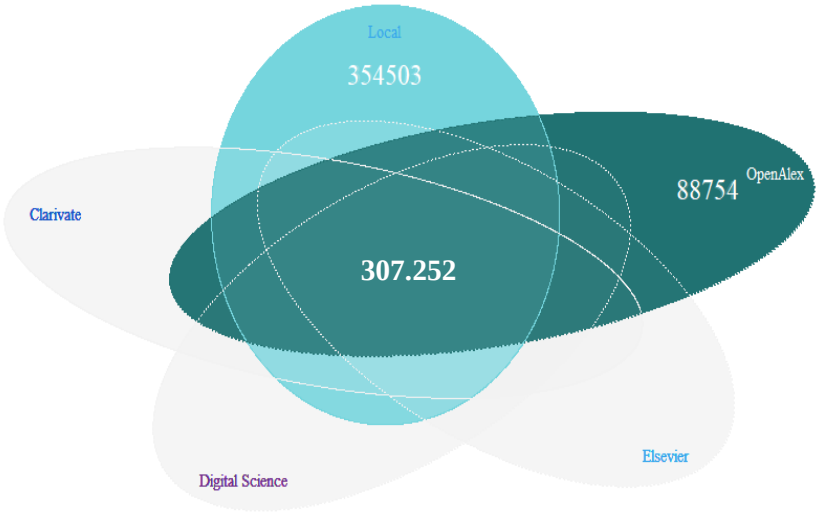
1b: Local versus Digital Science

The matching procedure identified **300.917** publications (23.147 per year on average) described in both **Local** and **Digital Science**, out of a total of 955.404. This represents **31,5%** of the total number of publications.

The matching procedure identified **318.053** publications (24.466 per year on average) described in both **Local** and **Elsevier**, out of a total of 955.404. This represents **33,3%** of the total number of publications.



1c: Local versus Elsevier



1d: Local versus OpenAlex

The matching procedure identified **307.252** publications (23.635 per year on average) described in both **Local** and **OpenAlex**, out of a total of 955.404. This represents **32,2%** of the total number of publications.

Table 2: Intersection, and unique publications for Local, Clarivate, Digital Science, Elsevier and OpenAlex

Coverage (2011-2023)	Intersection of 2				Only in 1				
	Loc+Cla	Loc+DS	Loc+Els	Loc+Opal	Local	Clarivate	Digital Science	Elsevier	OpenAlex
Total number	301.307	300.917	318.053	307.252	354.503	36.262	4.338	23.943	88.754
Average/year	23.177	23.147	24.466	23.635	27.269	2.789	334	1.842	6.827

The publications identified in the intersection between Local and the individual Global sources (Intersection of 2), comprise:

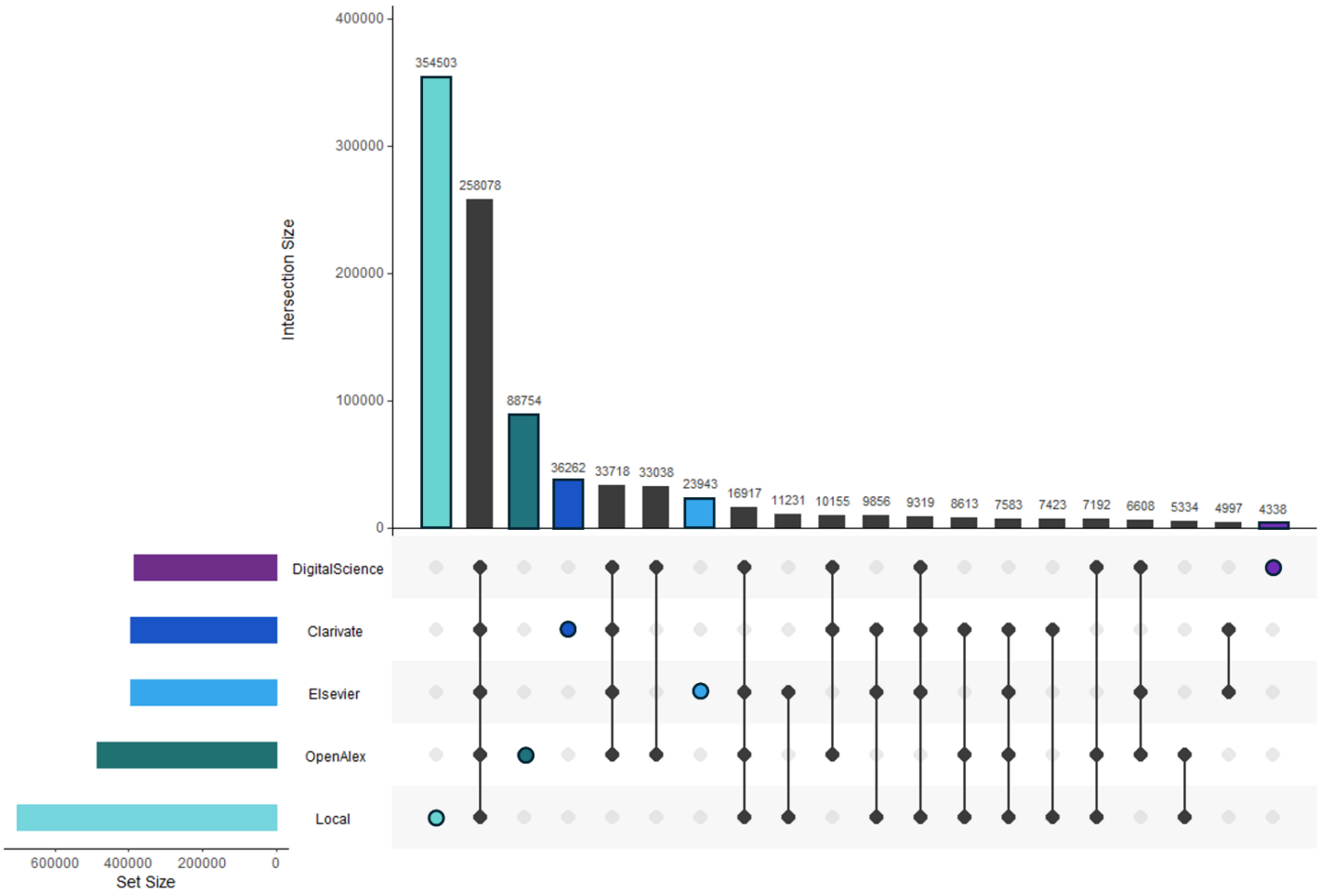
- 301.307 publications in **Local** as well as **Clarivate**.
  - **31,5%** of all publications identified (955.404).
  - **86,7%** of all publications in the intersection of **Local** and all **Global** sources (347.683).
    - Of these, **2,1%** (7.423) are unique for only **Local** and **Clarivate**.
- 300.917 publications in **Local** as well as **Digital Science**.
  - **31,5%** of all publications identified (955.404).
  - **86,5%** of all publications in the intersection of **Local** and all **Global** sources (347.683).
    - Of these, **0,2%** (633) are unique for only **Local** and **Digital Science**.
- 318.051 publications in **Local** as well as **Elsevier**.
  - **33,3%** of all publications identified (955.404).
  - **91,5%** of all publications in the intersection of **Local** and all **Global** sources (347.683).
    - Of these, **3,2%** (11.231) are unique for only **Local** and **Elsevier**.
- 307.252 publications in **Local** as well as **OpenAlex.**
  - **34,2%** of all publications identified (955.404).
  - **88,4%** of all publications in the intersection of **Local** and all **Global** sources (347.683).
    - Of these, **1,5%** (5.334) are unique for only **Local** and **OpenAlex**.

In **Figure 2** and **Table 3** the number of unique publications (Only in 1) from each of the data sources are presented.

Figure 2 visualizes the overlap between publication sets from Local and Global sources. Each row corresponds to a publication set, with horizontal bar charts representing the size of each set. Each column represents a possible intersection, corresponding to segments in a Venn diagram, while the bar charts at the top show the size of the intersections. The filled-in dots indicate which sets are part of an intersection. To read the figure, follow the dots in a column to identify the sets included in the corresponding intersection, and refer to the bar chart for its size.

The first column corresponds to the publications found only in **Local**. The green (third) column represents publications only found in **OpenAlex**, while the blue (fourth) column represents publications found only in **Clarivate**. The light blue (seventh) column indicates publications found only in **Elsevier**, and the last violet column presents publications found only in **Digital Science**.

Figure 2: Unique Publications in Local, Clarivate, Digital Science, Elsevier and OpenAlex



When comparing the number of unique publications from each data provider to the total number of publications received in total from **all** five data sources (955.404) the following shares of unique records are evident:

- **37,1%** (354.503) of all publications are received **only** from **Local**
- **9,3%** (88.754) of all publications are received **only** from **OpenAlex**
- **3,8%** (36.262) of all publications are received **only** from **Clarivate**

- **2,5%** (23.943) of all publications are received **only** from **Elsevier**
- **0,5%** (4.338) of all publications are received **only** from **Digital Science**

Table 3: Intersection, and unique publications for Local, Clarivate, Digital Science, Elsevier and OpenAlex

Coverage (2011-2023)	Intersection of 2				Only in 1				
	Loc+Cla	Loc+DS	Loc+Els	Loc+Opal	Local	Clarivate	Digital Science	Elsevier	OpenAlex
Total number	301.307	300.917	318.053	307.252	354.503	36.262	4.338	23.943	88.754
Average/year	23.177	23.147	24.466	23.635	27.269	2.789	334	1.842	6.827

Key observation

- Overall, the highest number of publications shared between the **Local** and **Global** sources is observed for **Local** and **Elsevier**, followed by **Local** and **OpenAlex**, with the lowest number found for **Local** and **Digital Science**.
- This may suggest a higher level of integration between **Local** and **Elsevier**. Specifically, PURE, an Elsevier product, is widely used by Danish research institutions, which often import a significant number of publications from Elsevier's databases to maintain their local CRIS.
- **Local** has the highest number of unique publications pertaining to the different scope and coverage than the Global sources.
- Of the Global data sources the highest number of unique publications is found in **OpenAlex,** indicating its broader coverage as a Global source. This will be analysed in more detail in the next section.
- In contrast, **Digital Science** has the lowest number of unique publications, which can likely be attributed to the large overlap with **OpenAlex**, encompassing 33.038 publications[13]. This overlap is potentially explained by the strong reliance of both **Digital Science** and **OpenAlex** on Crossref data, which **OpenAlex** clearly supplements with many other sources.

[13] For details, see the raw labelled "2 out of 4 (DS+Opal)" in Table A1 in the Appendix.

4. UNIQUE PUBLICATIONS

This section of the report will explore the unique publications (Only in 1) from each of the five data sources in greater detail, with the aim of understanding the specific contributions these data sources bring to the Research Portal Denmark publication databases.

To better understand and compare the unique publications from the five data sources, as well as the specific characteristics provided by each data provider, the following subsections examine various metadata elements relevant to bibliographic and bibliometric data sources. The elements compared across the four data sources, when available, include:

- Publication Types
- Languages
- Subject Classifications

4.1 PUBLICATION TYPES

Table 4a-4e presents the breakdown of publication types as defined and indexed by each of the five data providers, focusing on the unique publications from each source. Publication types and their definitions vary from provider to provider, hence, this overview of unique publications per publication type is separated into five tables, each pertaining to the publication type definition of the respective data source.

Table 4a: Local unique publications 2011-2023

Publication Types	Number	Share
Local		
Journal Article	61.808	17,4%
Book Chapter	54.860	15,5%
Conference Abstract	42.914	12,1%
Other	37.569	10,6%
Conference Paper	34.543	9,7%
Newspaper Article	24.122	6,8%
Report	21.980	6,2%
Conference Poster	18.411	5,2%
Book	17.594	5,0%
Thesis (Ph.D/Doctoral)	14.876	4,2%
Journal Book Review	6.856	1,9%
Working Paper/Preprint	6.484	1,8%
Book. Preface	5.339	1,5%
Report Chapter	3.267	0,9%
Journal Comment	2.825	0,8%
Journal Review	1.030	0,3%
Lecture Notes	680	0,2%
Total number and share of unique publications	**354.503**	**37,1%**

Note: Because of the analytical approach where multiple Local publication records are in some instances considered and counted as **one** publication, a Local publication may in some instances be assigned to more than one publication type. This is not the case for any of the Global publication records[14].

Table 4b: Clarivate unique publications 2011-2023

Publication Types	Number	Share
Clarivate		
Meeting Abstract	23.993	66,2%
Book Chapter	4.212	11,6%
Article	2.572	7,1%
Proceedings Paper	2.302	6,3%
Editorial Material	1.962	5,4%
Book Review	803	2,2%
Letter	114	0,3%
Review	114	0,3%
Correction/Retraction	88	0,2%
Other	33	0,1%
News Item	24	0,1%
Biographical-Item	30	0,1%
Book Review	13	0,04%
Book	2	0,01%
Total number and share of unique publications	**36.262**	**3,8%**

Table 4c: Digital Science unique publications 2011-2023

Publication Types	Number	Share
Digital Science		
Article	1.940	44,7%
Preprint	1.273	29,3%
Chapter	939	21,6%
Proceeding	157	3,6%
Monograph	29	0,7%
Total number and share of unique publications	**4.338**	**0,5%**

[14] The total count of publications distributed across all Local publication types is 355.158. Local publication types not affected are: Newspaper Articles, Thesis (Ph. D/Doctoral), Conference Posters Lecture Notes.

Table 4d: Elsevier unique publications 2011-2023

Publication Types	Number	Share
Elsevier		
Conference Paper	8.529	35,6%
Book Chapter	5.139	21,5%
Article	4.474	18,7%
Note	2.303	9,6%
Editorial	1.687	7,0%
Book	828	3,5%
Review	520	2,2%
Letter	316	1,3%
Erratum/Retratced	59	0,2%
Short Survey	58	0,2%
Conference Review	28	0,1%
Data Paper	1	0,004%
Abstract	1	0,004%
Total number and share of unique publications	**23.943**	**2,5%**

Table 4e: OpenAlex unique publications 2011-2023

Publication Types	Number	Share
OpenAlex		
Article	64.729	72,9%
Preprint	9.726	11,0%
Book Chapter	7.007	7,9%
Author Response	1.635	1,8%
Book	1.440	1,6%
Paratext	1.051	1,2%
Other	864	1,0%
Review	829	0,9%
Report	685	0,8%
Editorial	330	0,4%
Dissertation	195	0,2%
Letter	119	0,1%
Erratum	74	0,1%
Reference Entry	50	0,1%
Supplementary Materials	18	0,02%
Retraction	2	0,002%
Total number and share of unique publications	**88.754**	**9,3%**

Key observations

- **Local** indexes a broader spectrum of publication types than the global sources (e.g., Posters, Thesis' Reports, Newspaper Articles and Other), resulting in a high number of unique publications across many publication types. If looking at unique publications spread across some of the more common publication types – e.g. Journal Articles (17,4%), Book Chapters (15,5%), Conference Abstracts (12,1%), Conference Papers (9,7%) and Books (5,0%), these Publication Types accounts for more than half of the unique publications at 59,7% (211.719).
- **Clarivate's** largest share of unique publications is indexed as Meeting Abstracts, accounting for 66,2% (23.993) of all Clarivate's unique publications. Second comes Book Chapters with 11,6%.
- **Digital Science** operates with fewer publication types compared to the other sources. Besides Articles with a share of 44,7% (1.940), the largest share of unique publications is indexed as Preprints accounting for 29,3% (1.273) of all Digital Science's unique publications.
- **Elsevier's** largest share of unique publications is indexed as Conference Paper, accounting for 35,6% (8.529) of all Elsevier's unique publications. Second comes Book Chapters with 21,5%.
- **OpenAlex** has a predominant share of unique publications indexed as Articles, accounting for 72,9% (64.729) of all OpenAlex's unique publications. OpenAlex also accounts for a sizable share of unique Preprints at 11% (9.726) of all OpenAlex's unique publications.
- For all five data sources, Book Chapters constitute a significant portion of the unique publications: **Digital Science** 21,6% (939), **Elsevier** at 21,5% (5.139); **Local** 15,5% (54.860); **Clarivate** 11,6% (4.212) and **OpenAlex** 7,9% (7.007).

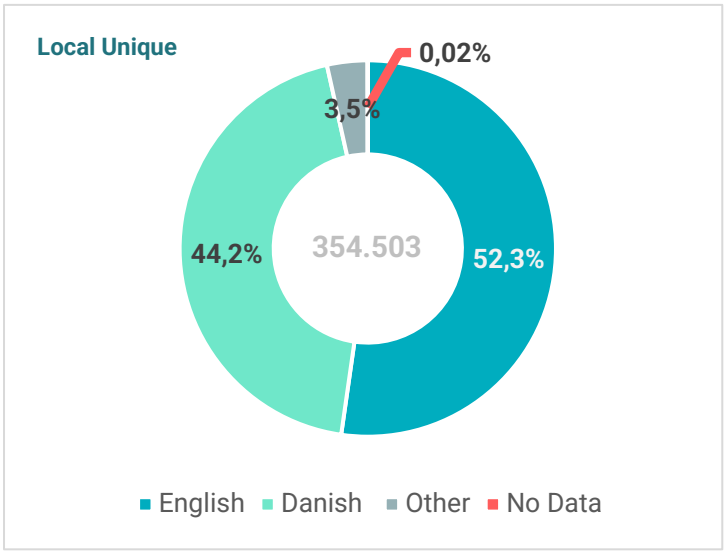
4.2 LANGUAGES

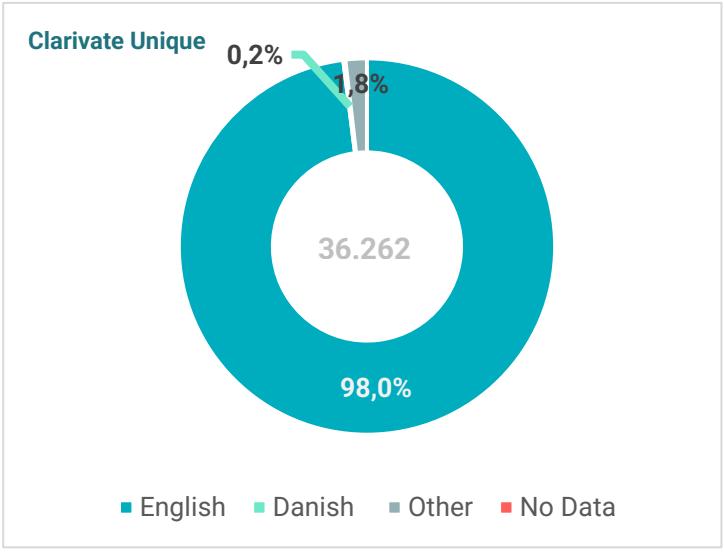
Figures 3a-3e presents overviews of the share of unique publications indexed with the language English, Danish or Other[15] for each of the five data sources. The share of publications without an indexed language is indicated as 'No Data'. In the centre of each figure, the total number of unique publications is displayed.

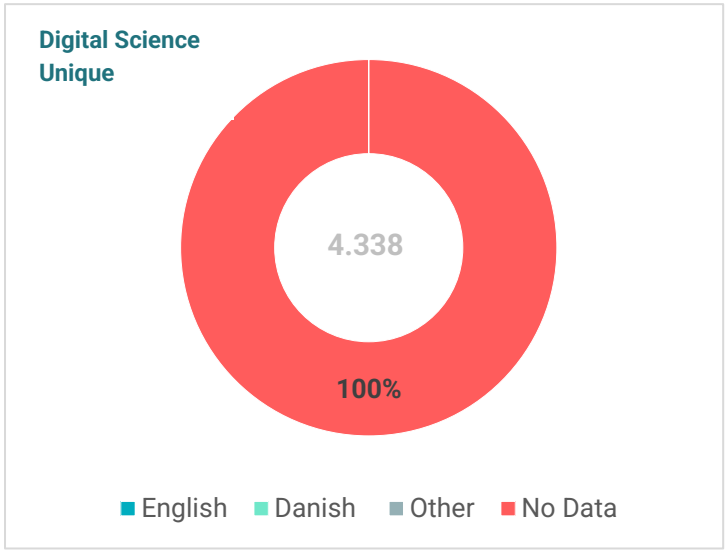
Table 5 gives an overview of the exact numbers behind the shares presented in the figures. A heatmap is used in the table to indicate the coverage of unique publications per language category **for each data source,** with varying shades from light to dark colours. If a data source does not classify language in their indexing, it is marked as N/A in the table.

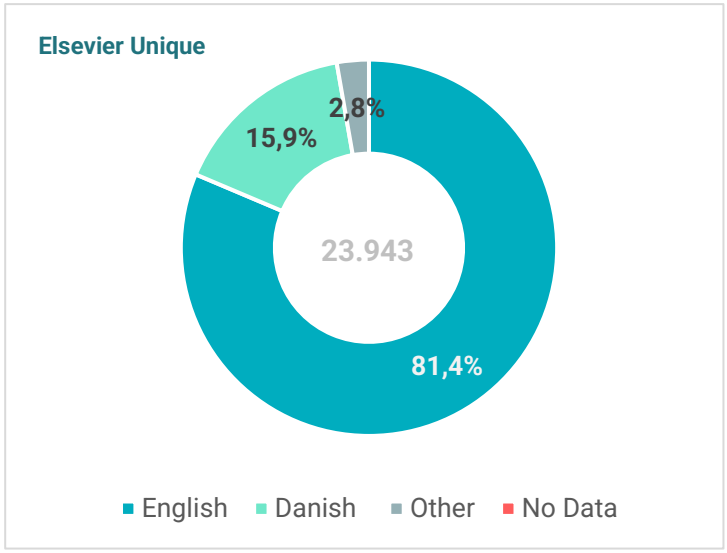
[15] The term 'Other' covers publications indexed as all other languages than Danish and English.

Figures 3a-e: Share of unique publications indexed as English, Danish or Other languages in the five data sources: Local, Clarivate, Digital Science, Elsevier and OpenAlex









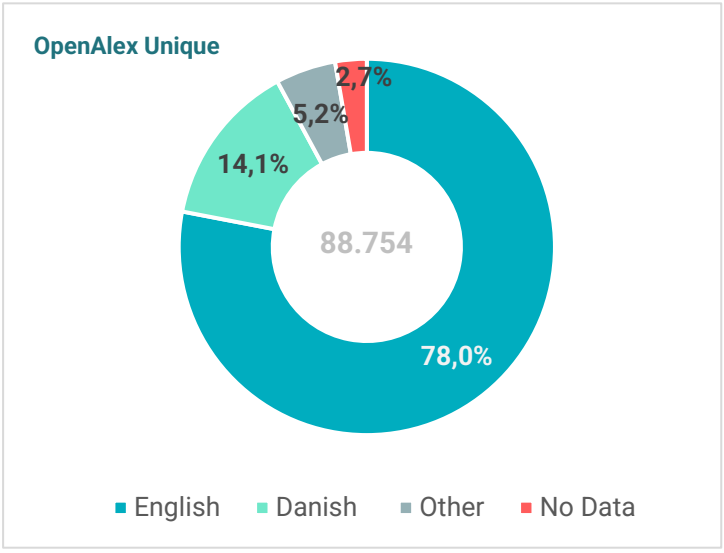


Table 5: Number of unique publications indexed as English, Danish or Other languages in the five data sources: Local, Clarivate, Digital Science, Elsevier and OpenAlex

Per data source: Lowest to highest number of unique publications

Language (2011-2023)	Only in 1				
	Local	Clarivate	Digital Science	Elsevier	OpenAlex
English	185.282	35.531	N/A	19.483	69.245
Danish	156.852	63	N/A	3.801	12.505
Other	12.321	668	N/A	659	4.599
No Data	61	0	4.338	0	2.405
Total number of unique publications	**354.503**	**36.262**	**4.338**	**23.943**	**88.754**

Key observations

- 44,2% (156.852) of all unique publications indexed in **Local** are published in Danish.
- In contrast, 15,9% (3.801) of the unique publications indexed in **Elsevier,** 14,1% (12.505) in **OpenAlex**[16] and only 0,2% (63) of the unique publications indexed in **Clarivate** are published in Danish.
- **Elsevier** and **OpenAlex** have a similar share of unique publications indexed as Danish - **OpenAlex** with 14,1% (12.505) and **Elsevier** 15,9% (3.801).
- **Digital Science** does not use -or index language as a metadata element in any of its publications. However, they do index Danish language publications. The absence of language indexing makes it impossible to determine the share of Danish language publications in Dimensions.

4.3 SUBJECT CLASSIFICATIONS

To facilitate the comparison of subject classifications, the Danish subject classification **DK Main Research Areas** is used. These areas include Science/Technology, Medical Science, Humanities, and Social Science. The subject classification was developed in connection with the discontinued Danish National Bibliometric Research Indicator (BFI) and is still used in the Danish Open Access Indicator[17]. It is the only subject classification, as of now, used across Research Portal Denmark databases, hence explaining the use of this classification to investigate the unique records further. For Local Data, the research areas are registered locally by the institutions on the publication records themselves, while for all four Global sources a mapping from their individual various subject classification has been used[18].

Figure 4a-4e provide overviews of the share of unique publications within the four **DK Main Research Areas** for each of the five data sources. The share of publications without any subject classification indexed is marked as 'No Data'. The total number of unique publications is shown at the centre of each figure.

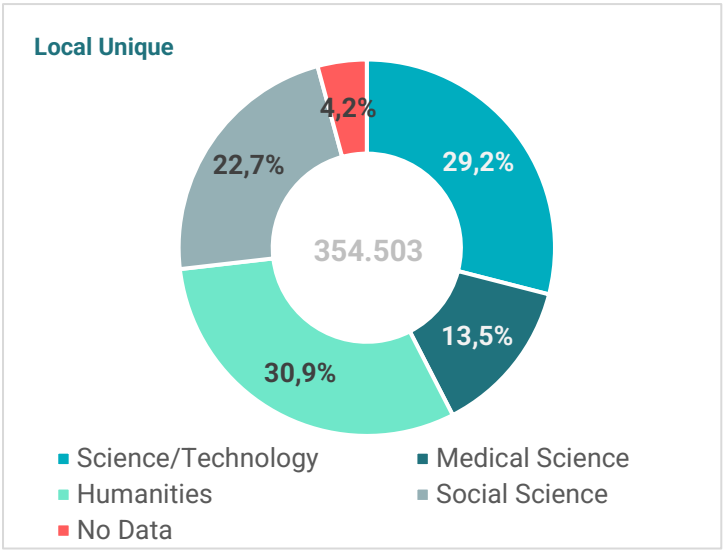
Table 6 gives an overview of the numbers behind the shares presented in the figures. A heatmap is used in the table to indicate the coverage of unique publications per subject classification category **for each data source,** with varying shades from light to dark colours.

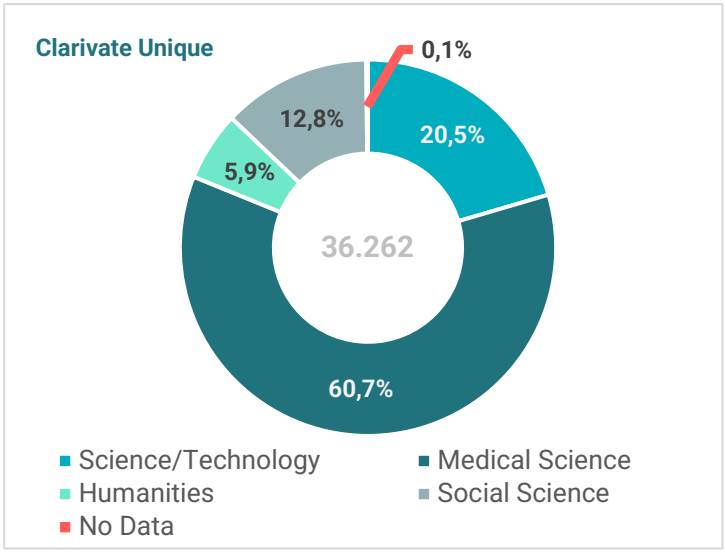
[16] Note that the numbers for Danish language publications in OpenAlex might increase due to the improvements in the quality of its language metadata. Recent assessment of the quality of language metadata shows that OpenAlex tends to overestimate the place of English and underestimate that of other languages.

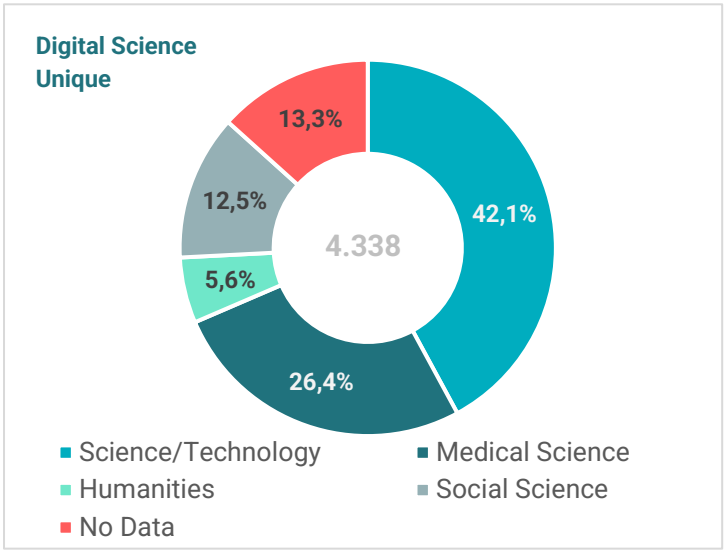
[17] The Danish Open Access Indicator is an analytical service under Research Portal Denmark - https://oaindikator.dk/en/

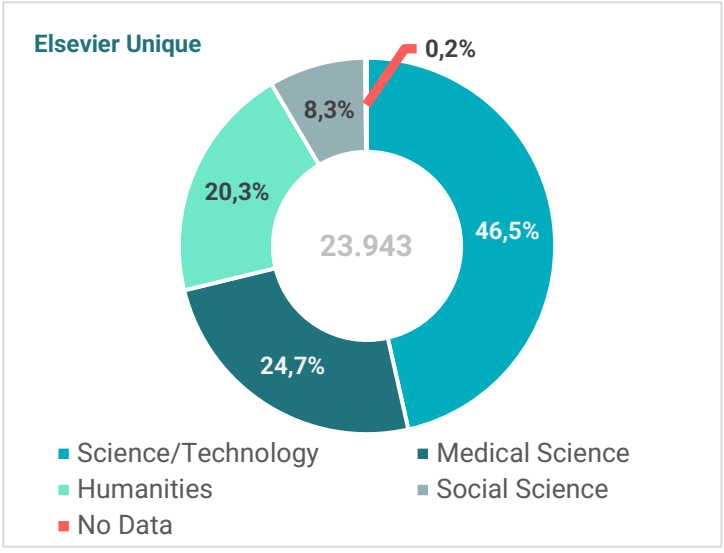
[18] For the full documentation describing the NORA subject classifications mappings read the technical documentation: https://docs.google.com/document/d/1JtDzj5f_s_IbbsiFb9CQtjMcXJzwew5-M07Ph_RTgJw/edit#heading=h.pq0xadoy8s7u

Figures 4a-e: Share of unique publications per DK Main Research Areas in the five data sources: Local, Clarivate, Digital Science, Elsevier and OpenAlex









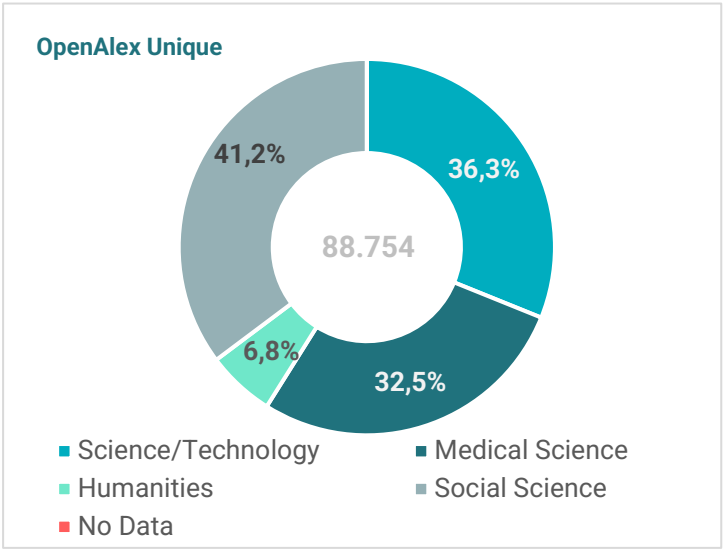


Table 6: Number of unique publications per DK Main Research Areas in the five data sources: Local, Clarivate, Digital Science, Elsevier and OpenAlex

Per data source: Lowest to highest number of unique publications

DK Main Research Areas (2011-2023)	Only in 1				
	Local	Clarivate	Digital Science	Elsevier	OpenAlex
Science/Technology	103.464	7.417	1.826	11.126	32.258
Medical Science	48.026	22.011	1.147	5.923	28.844
Humanities	109.682	2.154	245	4.862	6.055
Social Science	80.646	4.629	542	1.996	36.527
No Data	15.053	51	578	36	0
Total number of unique publications	**354.503**	**36.262**	**4.338**	**23.943**	**88.754**

Note: A single publication may be assigned to more than one DK Main Research Area[19].

Key observations

- **Local** has the greatest share of unique publications within the subject classification 'Humanities', with 30,9% (109.682), and the lowest within 'Medical Science' with 13,5% (48.026).
- **Clarivate** has the greatest share of unique publications within the subject classification 'Medical Science', at 60,7% (22.011), and the lowest within 'Humanities', at 5,9% (2.154).
- **Digital Science** has the greatest share of unique publications within the subject classification 'Science/Technology' at 42,1% (1.826), and the lowest within 'Humanities', at 5,6% (245).
- **Elsevier** has the greatest share of unique publications within the subject classification 'Science/Technology', at 46,5% (11.126), and the lowest within 'Social Science', at 8,3% (1.996).
- **OpenAlex** has the greatest share of unique publications within the subject classification 'Social Science', at 41,2% (36.527), and the lowest within 'Humanities', at 6,8% (6.055).

[19] This is resulting in a share total for Local of 100,7% and 116,8% for OpenAlex when combining the share numbers in Figure 4a and 4e.

5. BFI ANALYSIS

For an additional comparison and assessment of the coverage of the different data providers, publication data from the four Global data sources have been compared with the most recent Danish National Bibliometric Research Indicator (BFI) data set published in 2021 (based on publication data from 2020). The BFI 2021 data set includes the publication records that received BFI point based on a nationally curated channel list including journals, book series and conference proceedings with and without ISSN that were all considered as 'core' sources to publish Danish research within a given year.

As the BFI data set originally was created based on publication records harvested from the Danish universities' CRIS systems, its coverage would be expected to align 100% to Local Data, as all of these are currently Danish Data Providers for Research Portal Denmark.

The BFI analysis provides a general overview of BFI coverage across the four global sources, along with detailed insights into coverage at the Danish university level, offering a better understanding of the coverage.

5.1 BFI COVERAGE – AN OVERVIEW

Figure 5 presents the overall number and share of publications matched between the BFI data set and all- or the individual Global data sources respectively.

Table 7 gives an overview of the exact numbers and shares presented in the figures. In the table, a heatmap is used to indicate the coverage of unique publications **across the global data sources** (represented by varying shades from light to dark colours).

In **Appendix Table A4-A7**, the full BFI matching report is available, categorizing both matched and unmatched publications by Publication Types, DK Main Research Area, Language and Danish Data Providers/Universities.

Figure 5: Number and share of BFI 2021 publications matched with four Global data sources: Clarivate, Digital Science, Elsevier and OpenAlex

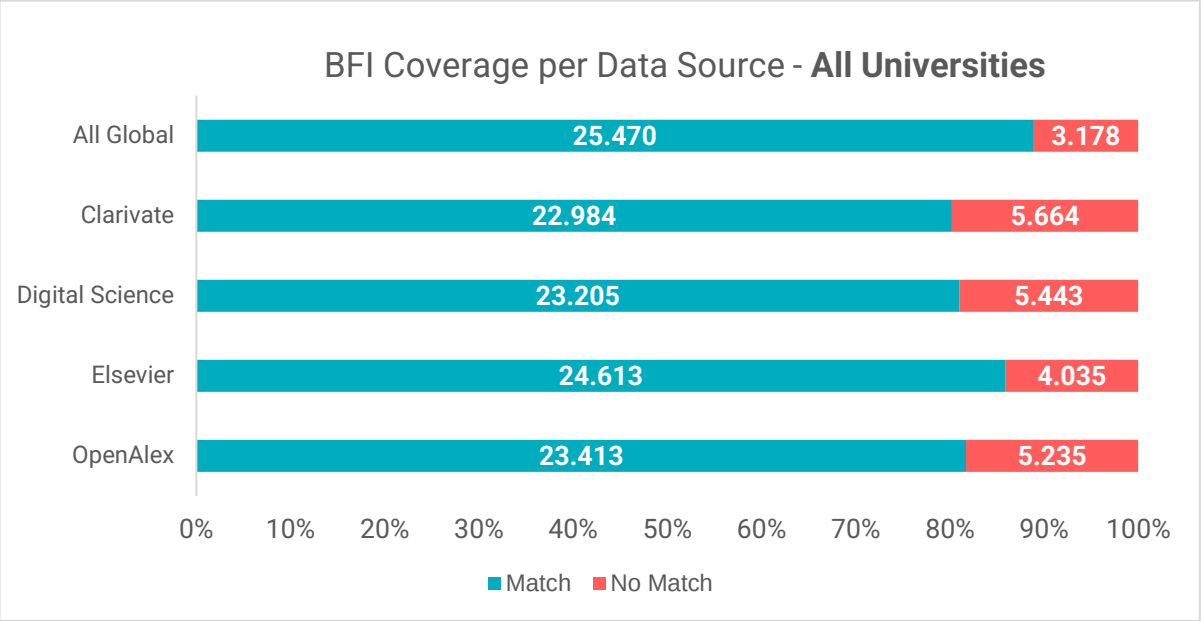


Table 7: Number and share of BFI 2021 publications matched with four Global data sources: Clarivate, Digital Science, Elsevier and OpenAlex

Across Global data sources: Lowest to highest number of matched or nonmatched BFI publications

All Universities	Match		No Match		Total
All Global	**25.470**	**88,9%**	**3.178**	**11,1%**	
Clarivate	22.984	80,2%	5.664	19,8%	
Digital Science	23.205	81,0%	5.443	19,0%	**28.648**
Elsevier	24.613	85,9%	4.035	14,1%	
OpenAlex	23.413	81,7%	5.235	18,3%	

Key observations

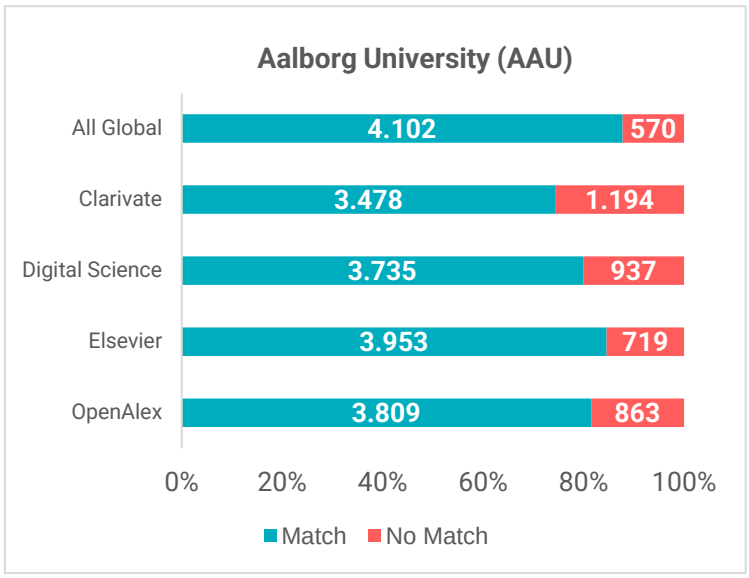
- When combining data from the four Global sources, the overall match percentage of the BFI 2021 dataset is notably high at 88,9% (25.470).
- Individually, three out of the four Global sources exhibit similar match percentages: **OpenAlex** at 81,7% (23.413); **Digital Science** at 81,0% (23.205) and **Clarivate** at 80,2% (22.984).
- While **Elsevier** exhibit a somewhat higher individual match rate at 85,9% (24.613).

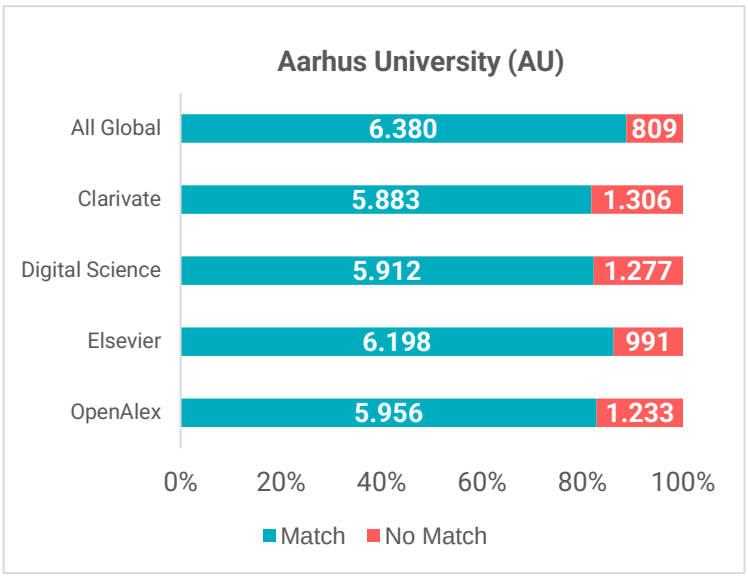
5.2 COVERAGE PER DANISH UNIVERSITY

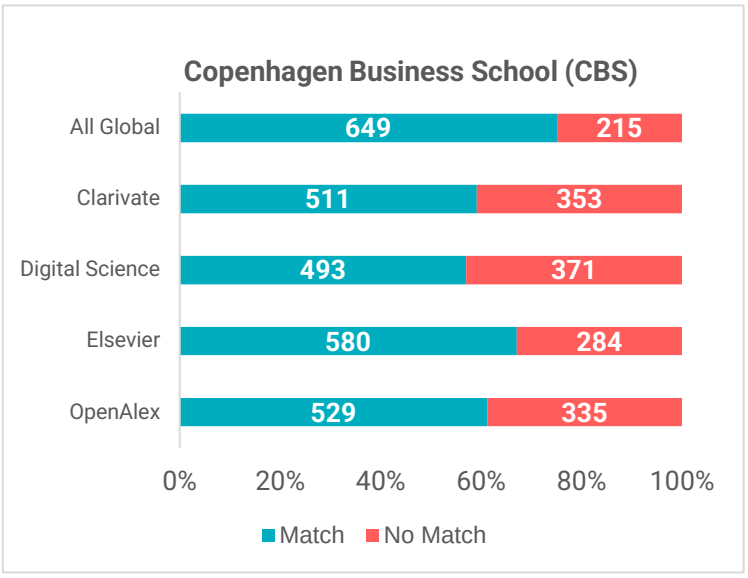
Figures 6a-6h presents the overall number and share of publications matched between the BFI data set and all- or the individual Global data sources respectively, displayed by Danish university.

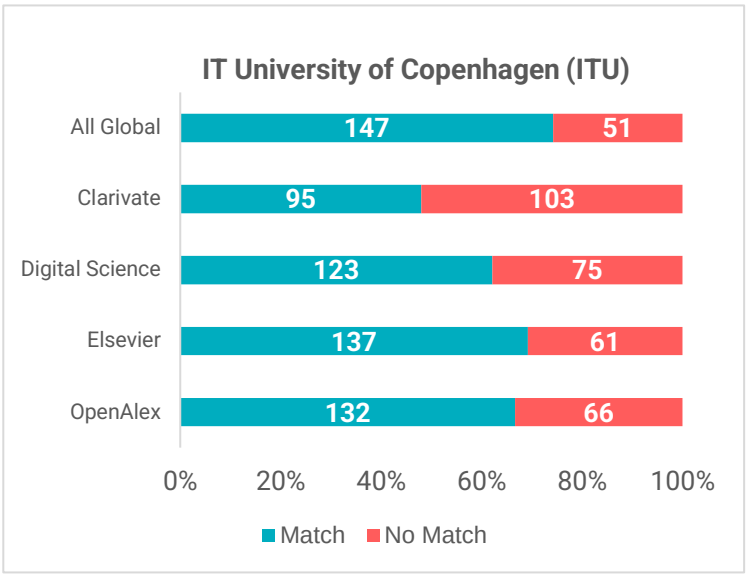
Tables 8a-8h gives a collected overview of the exact numbers and shares presented in the figures. In the table, a heatmap is used to indicate the coverage of unique publications **across the global data sources** (represented by varying shades from light to dark colours).

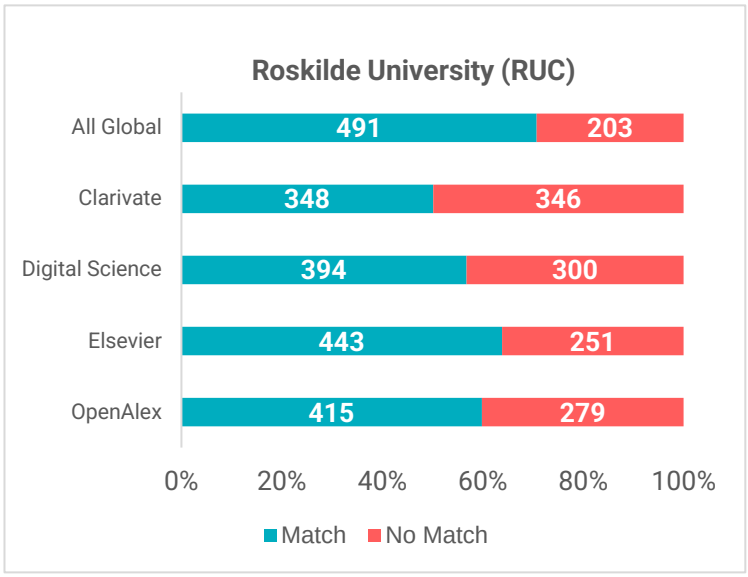
Figures 6a-h: Number and share of BFI 2021 publications matched with the four Global data sources: Clarivate, Digital Science, Elsevier and OpenAlex by Danish university

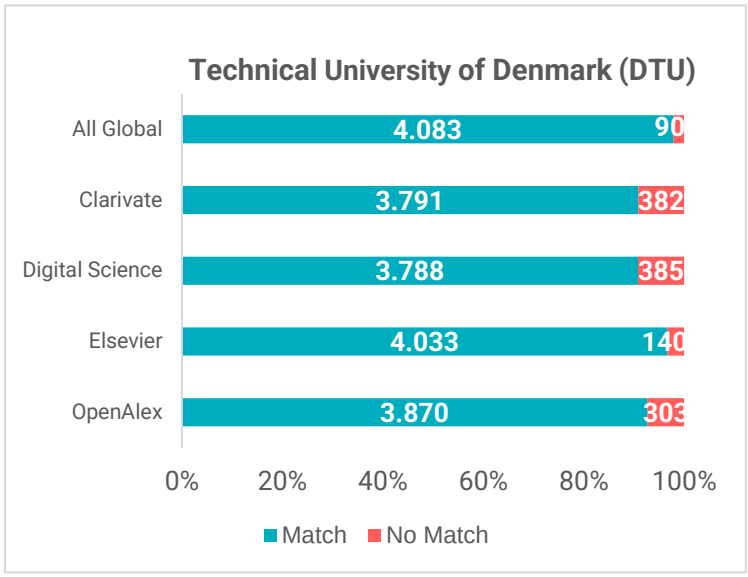


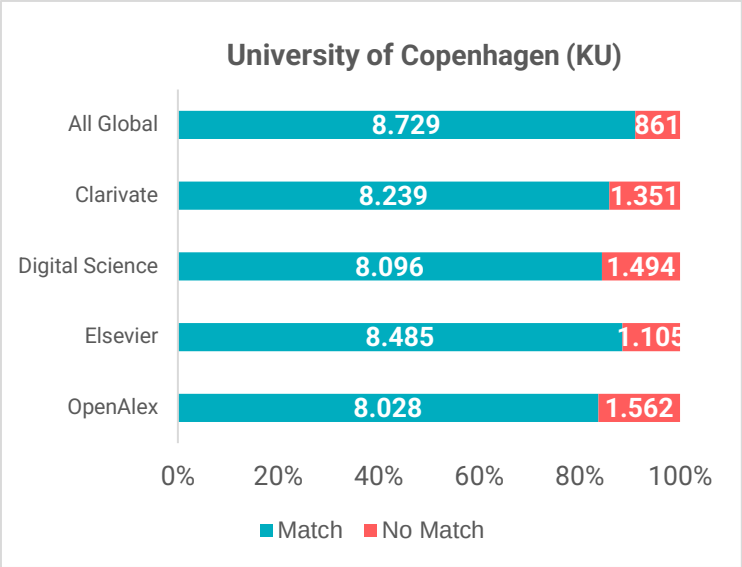


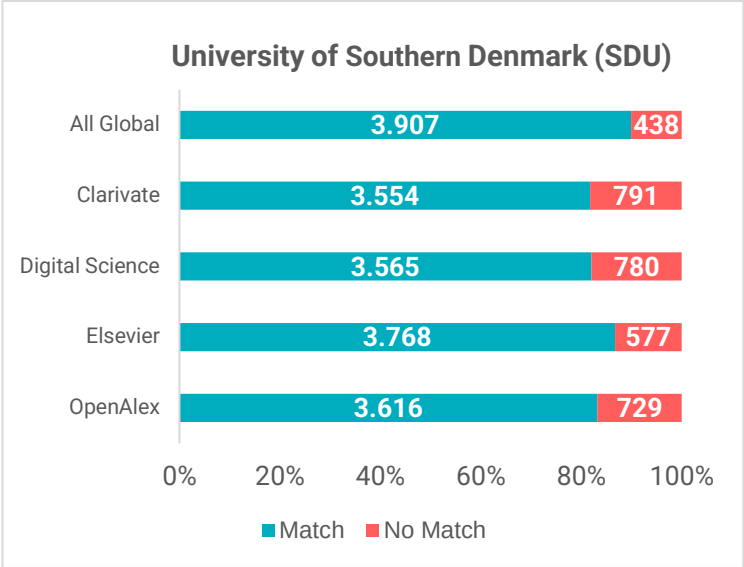












Tables 8a-h: Number and share of BFI 2021 publications matched with the four Global data sources: Clarivate, Digital Science, Elsevier and OpenAlex by Danish university

Across Global data sources: Lowest to highest number of matched or nonmatched BFI publications

AAU	Match		No Match		Total
All Global	4.102	87,8%	570	12,2%	
Clarivate	3.478	74,4%	1.194	25,6%	
Digital Science	3.735	79,9%	937	20,1%	4.672
Elsevier	3.953	84,6%	719	15,4%	
OpenAlex	3.809	81,5%	863	18,5%	

AU	Match		No Match		Total
All Global	6.380	88,7%	809	11,3%	
Clarivate	5.883	81,8%	1.306	18,2%	
Digital Science	5.912	82,2%	1.277	17,8%	7.189
Elsevier	6.198	86,2%	991	13,8%	
OpenAlex	5.956	82,8%	1.233	17,2%	

CBS	Match		No Match		Total
All Global	649	75,1%	215	24,9%	
Clarivate	511	59,1%	353	40,9%	
Digital Science	493	57,1%	371	42,9%	864
Elsevier	580	67,1%	284	32,9%	
OpenAlex	529	61,2%	335	38,8%	

ITU	Match		No Match		Total
All Global	147	74,2%	51	25,8%	
Clarivate	95	48,0%	103	52,0%	
Digital Science	123	62,1%	75	37,9%	198
Elsevier	137	69,2%	61	30,8%	
OpenAlex	132	66,7%	66	33,3%	

RUC	Match		No Match		Total
All Global	491	70,7%	203	29,3%	
Clarivate	348	50,1%	346	49,9%	
Digital Science	394	56,8%	300	43,2%	694
Elsevier	443	63,8%	251	36,2%	
OpenAlex	415	59,8%	279	40,2%	

DTU	Match		No Match		Total
All Global	4.083	97,8%	90	2,2%	
Clarivate	3.791	90,8%	382	9,2%	
Digital Science	3.788	90,8%	385	9,2%	4.173
Elsevier	4.033	96,6%	140	3,4%	
OpenAlex	3.870	92,7%	303	7,3%	

KU	Match		No Match		Total
All Global	8.729	91,0%	861	9,0%	
Clarivate	8.239	85,9%	1.351	14,1%	
Digital Science	8.096	84,4%	1.494	15,6%	9.590
Elsevier	8.485	88,5%	1.105	11,5%	
OpenAlex	8.028	83,7%	1.562	16,3%	

SDU	Match		No Match		Total
All Global	3.907	89,9%	438	10,1%	
Clarivate	3.554	81,8%	791	18,2%	
Digital Science	3.565	82,0%	780	18,0%	4.345
Elsevier	3.768	86,7%	577	13,3%	
OpenAlex	3.616	83,2%	729	16,8%	

Key observations

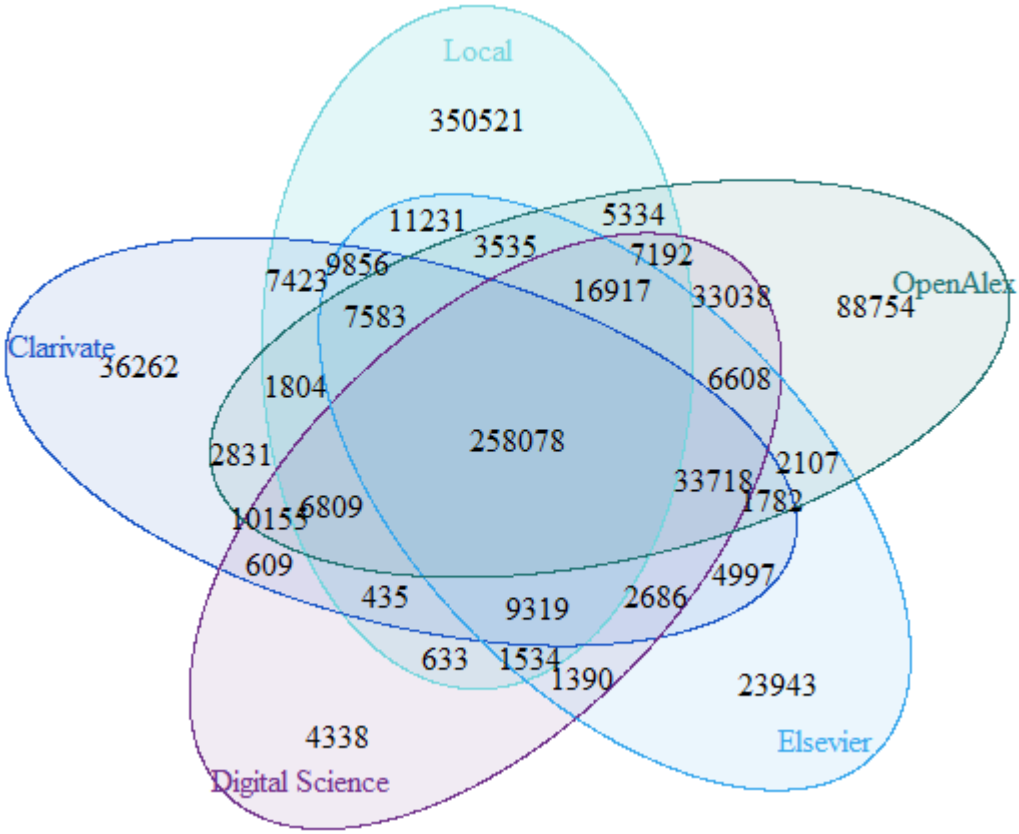
- Overall, the BFI coverage varies by university across individual Global data sources. However, a consistent pattern emerges, with **Elsevier** showing the highest coverage percentage across all cases.
- The BFI coverage for larger and multi-faculty Danish universities is generally more extensive and similar compared to smaller or more discipline-specific universities. For instance, the University of Copenhagen (KU) has a match percentage of 91,0% (8.729), the University of Southern Denmark (SDU) 89,9% (3.907), Aarhus University (AU); 88,7% (6.380) and Aalborg University (AAU) 87,8% (4.102) when considering all global data sources combined.
- Conversely, the BFI coverage for smaller or more discipline-specific Danish universities tends to be lower. For example, Copenhagen Business School (CBS) has a match percentage of 75,1% (649), IT University of Copenhagen (ITU) 74,2% (147) and Roskilde University (RUC) 70,7% (491) when considering all global data sources combined.
- An exception to these observations is Technical University of Denmark (DTU) which has the highest BFI coverage among all universities, with a match percentage of 97,8% (3.791) for all global data sources combined. This is complimented by the numbers in **Appendix Table A5** distributing BFI matches per DK Main Research Areas, where it is evident that publications within the classification 'Science/Technology' and in the publication channels used by DTU are very often covered by the global sources.
- Universities with a specific focus on humanities and/or social science (e.g. RUC and CBS) tend to have a lower BFI match percentage, than universities with a broader focus. This is complimented by the numbers in **Appendix Table A5**, where the number of publications within the classifications 'Humanities' and 'Social Sciences' are the ones with the lowest number of BFI matches.
- Universities that publish more extensively in Danish language publication channels (e.g. RUC and CBS) tend to have a lower BFI match percentage, compared to universities publishing predominantly in English language channels. **Appendix Table A6** supports this, showing that Danish publications have fewer BFI matches than English publications, with a BFI match percentage of 93,9% (25.020) for English publications, while only 20,5% (373) for Danish publications.

6. APPENDIX

6.1 APPENDIX 1 – COVERAGE ANALYSIS

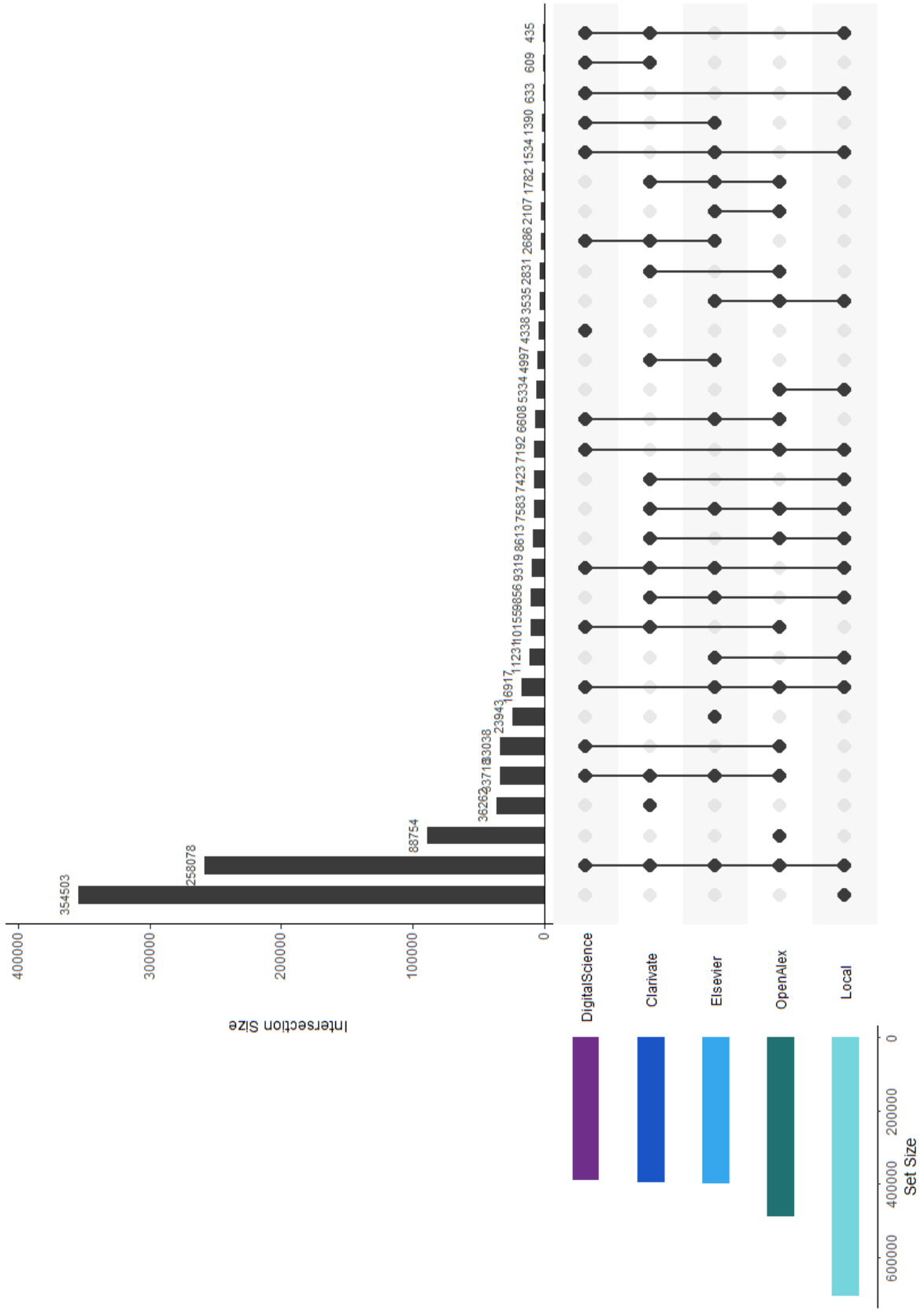
A total of 2.356.537 publication records were submitted to the matching procedure described in section 2, identifying 955.404 publications identified by one, two, three, four or all five sources (Union of all sources). This total number of publications is calculated as the sum of all the Venn diagram zones described in **Appendix Figure A1**.

Appendix Figure A1: Publications identified in all five sources: Local, Clarivate, Digital Science, Elsevier and OpenAlex



Appendix Figure A2 visualizes the overlap between publication sets from Local and Global sources. Each row corresponds to a publication set, with horizontal bar charts representing the size of each set. Each column represents a possible intersection, corresponding to segments in a Venn diagram (as illustrated above), while the bar charts at the top show the size of the intersections. The filled-in dots indicate which sets are part of an intersection. To read the figure, follow the dots in a column to identify the sets included in the corresponding intersection, and refer to the bar chart for its size. The first column corresponds to the publications found only in **Local**. The second column shows publications that are shared by all five sources, while the fifth column represents publications that are shared by all **Global** sources.

Appendix Figure A2: Publications identified in and across all data sources: Local, Clarivate, Digital Science, Elsevier and OpenAlex



Appendix Table A3 presents summary figures for the publications identified in and across all data sources.

Among the 955.404 identified publications, there are differences across the data sources:

- **Local** included 698.204 publication records (73,1% of the total, 53.708 per year on average).
- **Clarivate** included 389.804 publication records (40,8% of the total, 29.985 per year on average).
- **Digital Science** included 392.360 publication records (41,1% of the total, 30.182 per year on average).
- **Elsevier** included 390.334 publication records (40,9% of the total, 30.026 per year on average).
- **OpenAlex** included 485.835 publication records (50,9% of the total, 37.372 per year on average).

Appendix Table A3: Publications identified in and across all data sources: Local, Clarivate, Digital Science, Elsevier and OpenAlex

Coverage (2011-2023)		Total number	Percentage	Average/year
Total number	Local	698.204	80,6%	53.708
	Clarivate	389.804	45,0%	29.985
	Digital Science	392.360	45,3%	30.182
	Elsevier	390.334	45,0%	30.026
	OpenAlex	485.835	56,1%	37.372
Union		866.650	100,0%	66.665
All 5	Loc+Cla+DS+Els+Opal	258.078	29,8%	19.852
4 out of 5	Loc+Cla+DS+Els	9.319	1,1%	717
	Loc+Cla+DS+Opal	6.809	0,8%	524
	Loc+Cla+Els+Opal	7.583	0,9%	583
	Loc+DS+Els+Opal	16.917	2,0%	1.301
	Cla+DS+Els+Opal	33.718	3,9%	2.594
3 out of 4	Loc+Cla+DS	435	0,1%	33
	Loc+DS+Els	1.534	0,2%	118
	Loc+Cla+Els	9.856	1,1%	758
	Cla+DS+Els	2.686	0,3%	207
	Loc+Cla+Opal	1.804	0,2%	139
	Loc+DS+Opal	7.192	0,8%	553
	Loc+Els+Opal	3.535	0,4%	272
	Cla+DS+Opal	10.155	1,2%	781
	DS+Els+Opal	6.608	0,8%	508
	Cla+Els+Opal	1.782	0,2%	137
2 out of 4	Loc+Cla	7.423	0,9%	571
	Loc+DS	633	0,1%	49
	Loc+Els	11.231	1,3%	864
	Cla+DS	609	0,1%	47
	Cla+Els	4.997	0,6%	384
	DS+Els	1.390	0,2%	107
	Loc+Opal	5.334	0,6%	410
	Cla+Opal	2.831	0,3%	218
	Els+Opal	2.107	0,2%	162
	DS+Opal	33.038	3,8%	2.541
Only in 1	Local	354.503	40,9%	27.269
	Clarivate	36.262	4,2%	2.789
	Digital Science	4.338	0,5%	334
	Elsevier	23.943	2,8%	1.842
	OpenAlex	88.754	10,2%	6.827

6.2 APPENDIX 2 – BFI ANALYSIS

Appendix Table A4-A7gives an overview of the full BFI analysis of BFI record matches for the four Global data sources: Clarivate, Digital Science, Elsevier, OpenAlex. The analysis is based on the BFI 2021 dataset **AnnualReport-2020.zip**, harvested on the 30th of November 2021 and November 2024 Local Data. The full dataset consists of 28.764 publications (31.861 records) and of these 64 publications (80 records) and 52 patents (52 records) did not match with Local Data. Leaving 28.648 publications (31.729 records) for further matching with the four global data sources.

Appendix Table A4: Full BFI Analysis by Publication Type as indexed in Local Data

Publication Type	Match					No Match				
	All Global	Clarivate	Digital Science	Elsevier	OpenAlex	All Global	Clarivate	Digital Science	Elsevier	OpenAlex
Book	79	3	32	53	47	225	301	272	251	257
Book Chapter	1118	390	603	819	646	1660	2388	2175	1959	2132
Book Preface, Encycl. Entry	1			1			1	1		1
Conference Abstract	14	5	11	9	12	2	11	5	7	4
Conference Paper	2448	1600	1986	2375	2155	240	1088	702	313	533
Journal Article	20566	19763	19366	20117	19374	998	1801	2198	1447	2190
Journal Comment	5	5	5	5	4	1	1	1	1	2
Journal Review	1237	1217	1201	1234	1173	19	39	55	22	83
Other	1	1	1		1	1	1	1	1	1
Report	1				1	15	16	16	16	15
Report Chapter						4	4	4	4	4
Thesis (Doctoral)						13	13	13	13	13
Total	**25470**	**22984**	**23205**	**24613**	**23413**	**3178**	**5664**	**5443**	**4035**	**5235**

Appendix Table A5: Full BFI Analysis by DK Main Research Area as indexed in Local Data

DK Main Research Area	Match					No Match				
	All Global	Clarivate	Digital Science	Elsevier	OpenAlex	All Global	Clarivate	Digital Science	Elsevier	OpenAlex
Humanities	1506	941	1054	1247	1154	1496	2061	1948	1755	1848
Medical Science	9311	8937	8828	9163	8777	268	642	751	416	802
Science/Technology	12679	11613	11826	12449	11881	429	1495	1282	659	1227
Social Science	2432	1923	1933	2204	2050	1003	1512	1502	1231	1385
Total	**25470**	**22984**	**23205**	**24613**	**23413**	**3178**	**5664**	**5443**	**4035**	**5235**

Appendix Table A6: Full BFI Analysis by Language as indexed in Local Data

Language	Match					No Match				
	All Global	Clarivate	Digital Science	Elsevier	OpenAlex	All Global	Clarivate	Digital Science	Elsevier	OpenAlex
Danish	373	57	124	248	206	1450	1766	1699	1575	1617
Dutch						1	1	1	1	1
English	25020	22881	23026	24310	23146	1623	3762	3617	2333	3497
Finnish						1	1	1	1	1
French	8	3	3	5	6	14	19	19	17	16
German	24	8	15	13	16	55	71	64	66	63
Icelandic						1	1	1	1	1
Italian						2	2	2	2	2
Japanese	1	1	1	1	1	1	1	1	1	1
Norwegian	1			1		12	13	13	12	13
Portuguese	2	1	1	1	1		1	1	1	1
Spanish	11	7	9	7	11	5	9	7	9	5
Swedish	2		1	1		12	14	13	13	14
Undetermined	28	26	25	26	26	1	3	4	3	3
Total	**25470**	**22984**	**23205**	**24613**	**23413**	**3178**	**5664**	**5443**	**4035**	**5235**

Appendix Table A7: Full BFI Analysis by Data Provider / University

Data Provider / University	Match					No Match				
	All Global	Clarivate	Digital Science	Elsevier	OpenAlex	All Global	Clarivate	Digital Science	Elsevier	OpenAlex
AAU	4102	3478	3735	3953	3809	570	1194	937	719	863
AU	6380	5883	5912	6198	5956	809	1306	1277	991	1233
CBS	649	511	493	580	529	215	353	371	284	335
DTU	4083	3791	3788	4033	3870	90	382	385	140	303
ITU	147	95	123	137	132	51	103	75	61	66
KU	8729	8239	8096	8485	8028	861	1351	1494	1105	1562
RUC	491	348	394	443	415	203	346	300	251	279
SDU	3907	3554	3565	3768	3616	438	791	780	577	729
Total	**25470**	**22984**	**23205**	**24613**	**23413**	**3178**	**5664**	**5443**	**4035**	**5235**