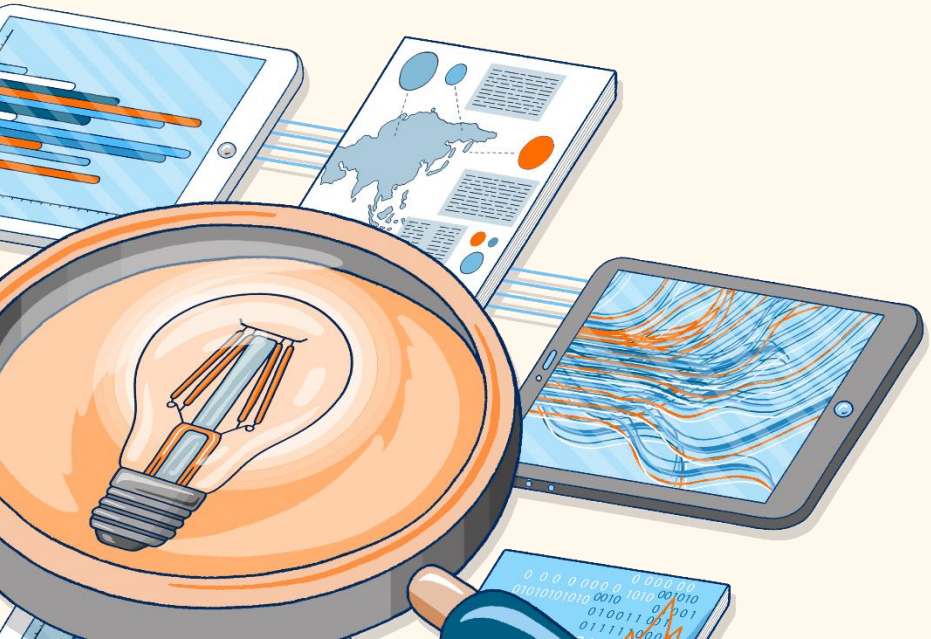
ELSEVIER

Pilot analysis of CRediT data: Exploring data coverage and role attribution patterns

Analytical Services

Introduction



Analytical Services

What is CRediT?

CRediT (Contributor Roles Taxonomy) is an initiative to acknowledge individual contributions to scholarly research. It was introduced by a collaborative group of publishers, funders, and institutions.

Developed: in 2012 following a workshop hosted by Harvard University and the Wellcome Trust to identify ways in which to provide visibility and recognition to scholarly contributions.

Adopted: Widely adopted in the following years, with increasing implementation in scholarly publishing.

The taxonomy: roles

CRediT offers a high-level taxonomy, consisting of 14 roles, that reflects the diverse contributions individuals can make to a scholarly output. These roles are designed to be applicable across all fields of scholarly endeavor. **Please refer to the annex for role descriptions.**

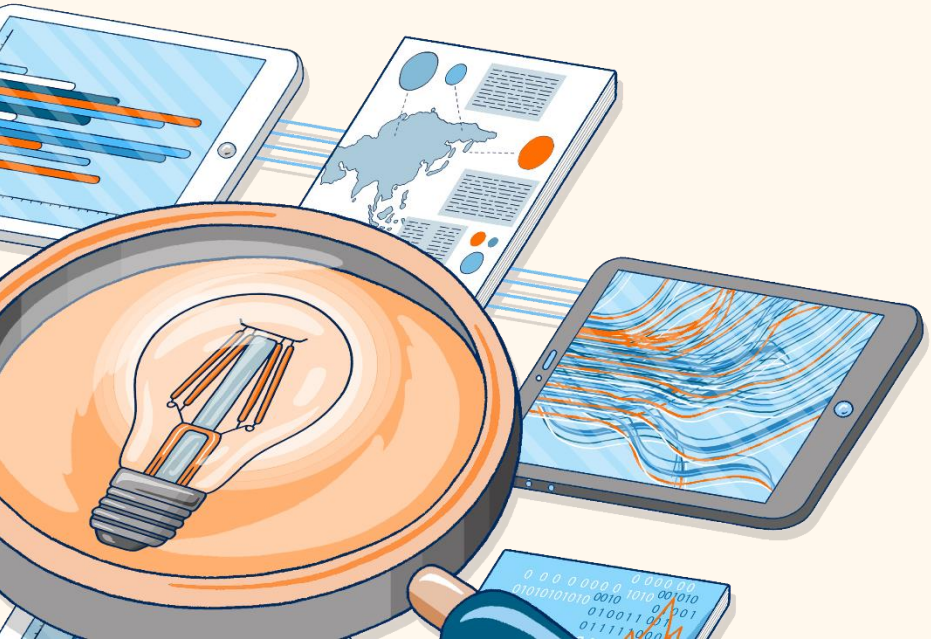
The taxonomy: implementation

Multiple Roles Possible – Individual contributors can be assigned multiple roles, and a given role can be assigned to multiple contributors;

Degree of Contribution Optional – Where multiple individuals serve in the same role, the degree of contribution can optionally be specified as ‘lead’, ‘equal’, or ‘supporting’;

Shared Responsibility – Corresponding authors should assume responsibility for role assignment, and all contributors should be given the opportunity to review and confirm assigned roles.

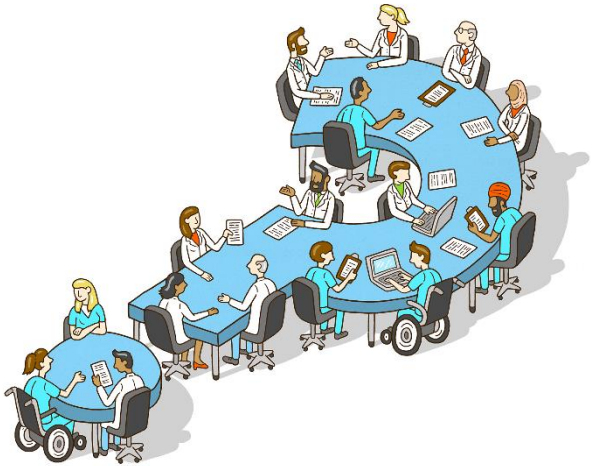
Analysis



Purpose of this pilot analysis

This analysis is an initial exploration of the CRediT data coverage and patterns associated with CRediT role attributions.

- The analysis covers the period of 2013-2022 with a focus on 2020-2022 when there is a significant increase in coverage.
- The country-level comparisons focus on the Netherlands, Belgium, Switzerland, the United Kingdom, Sweden, Denmark, Germany and Denmark.
- For subject-level analysis, the All Science Journal Classification (ASJC) is used.

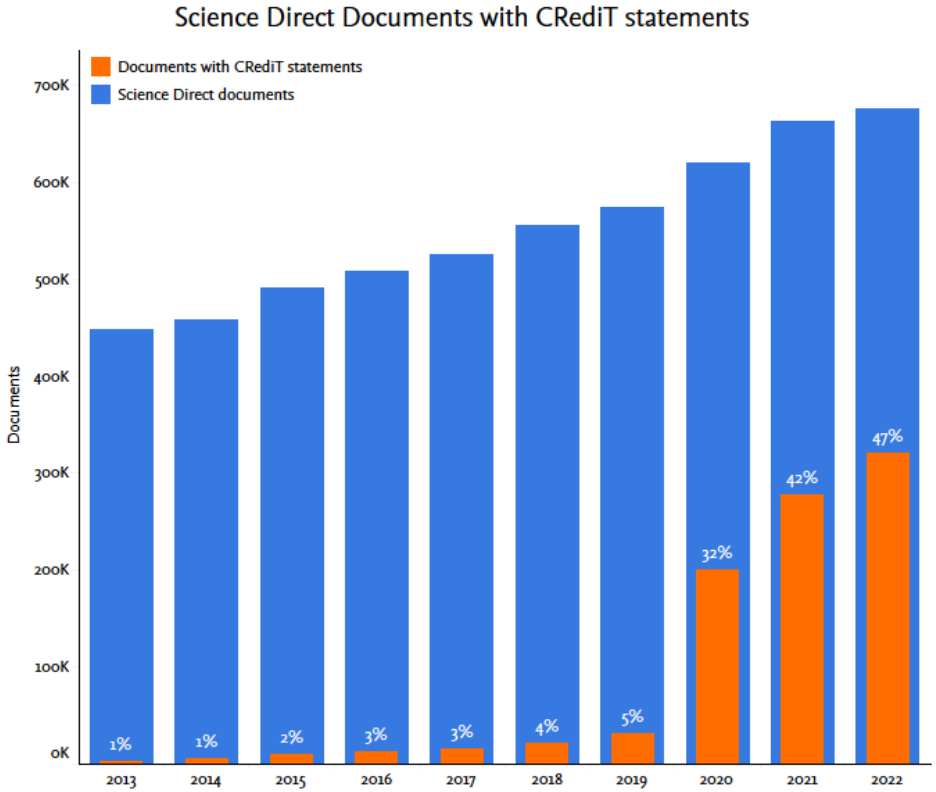


CRediT statements in Science Direct data

The number of Science Direct publications with CRediT statements has been gradually increasing since the development of the taxonomy in 2012.

From 2020, the share of publications with CRediT statements experienced a significant increase up to 47% in 2022.

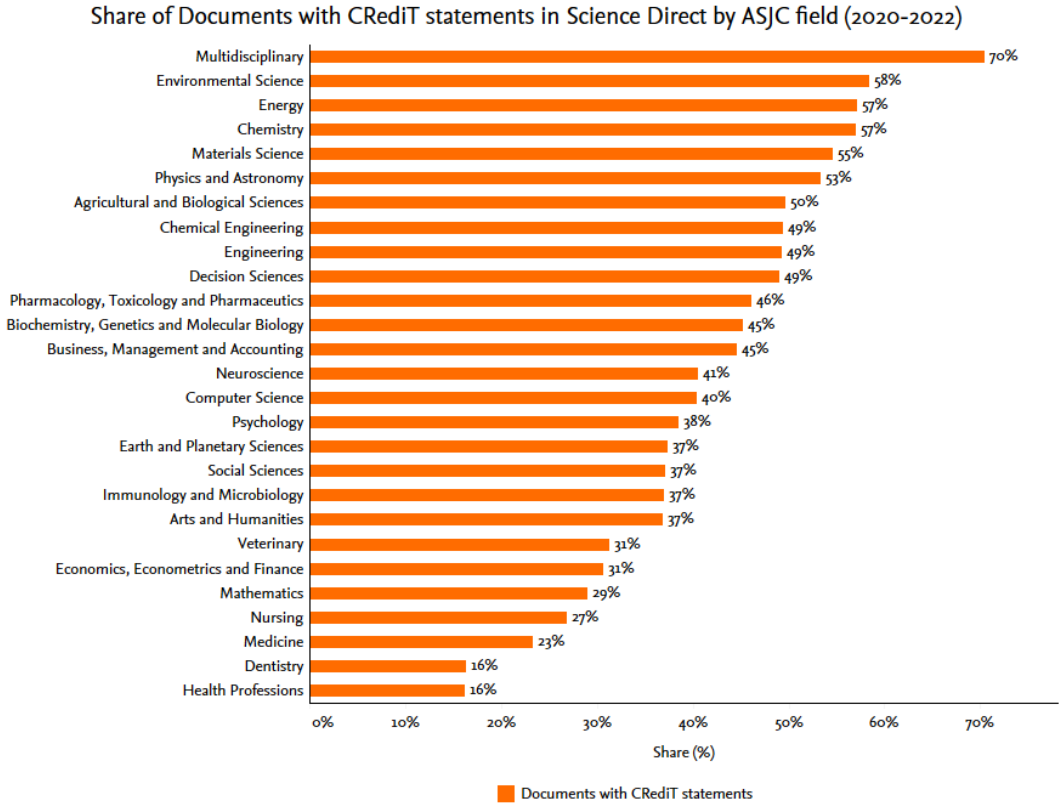
The figures likely indicate wider adoption among the publishers, journals and academics in the recent years.



Share of documents with CRediT by ASJC

The share of Science Direct documents with CRediT statements varies across subject fields.

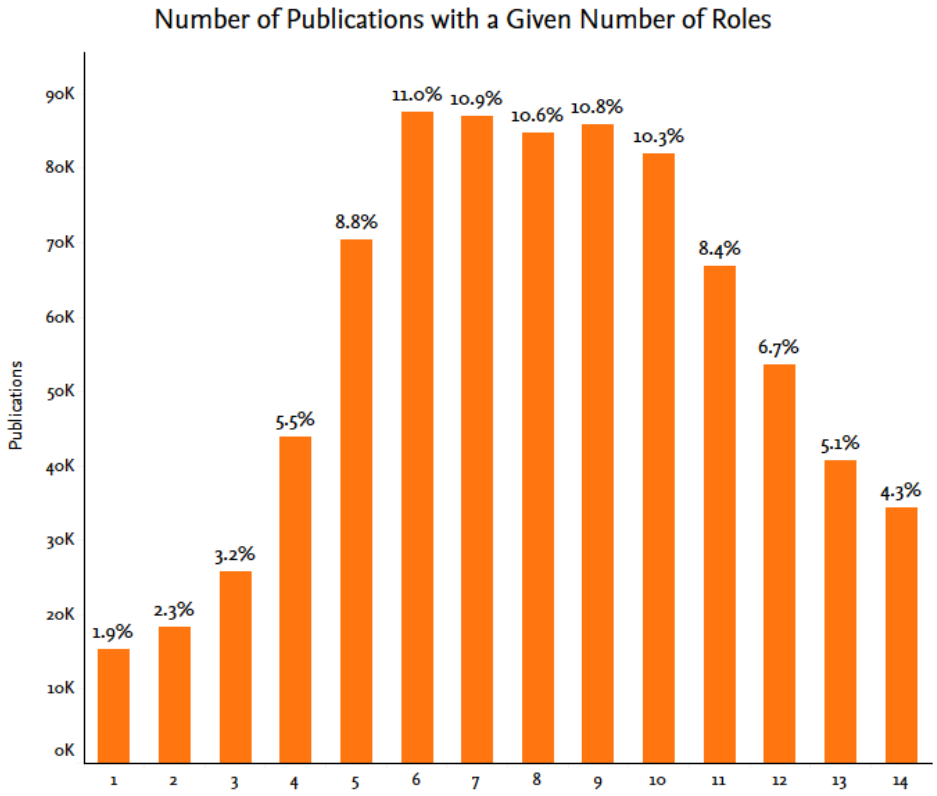
The Multidisciplinary category of the ASJC classification has the highest share of documents with CRediT data (70%). This likely stems from this category containing multiple flagship journals with a high level of CRediT adoption and robust measures in place to collect CRediT metadata.



Number of roles in publications

Publications differ in the number of roles that authors assume. The distribution of publications by number of roles is somewhat bell-shaped, but with a skewness to the right.

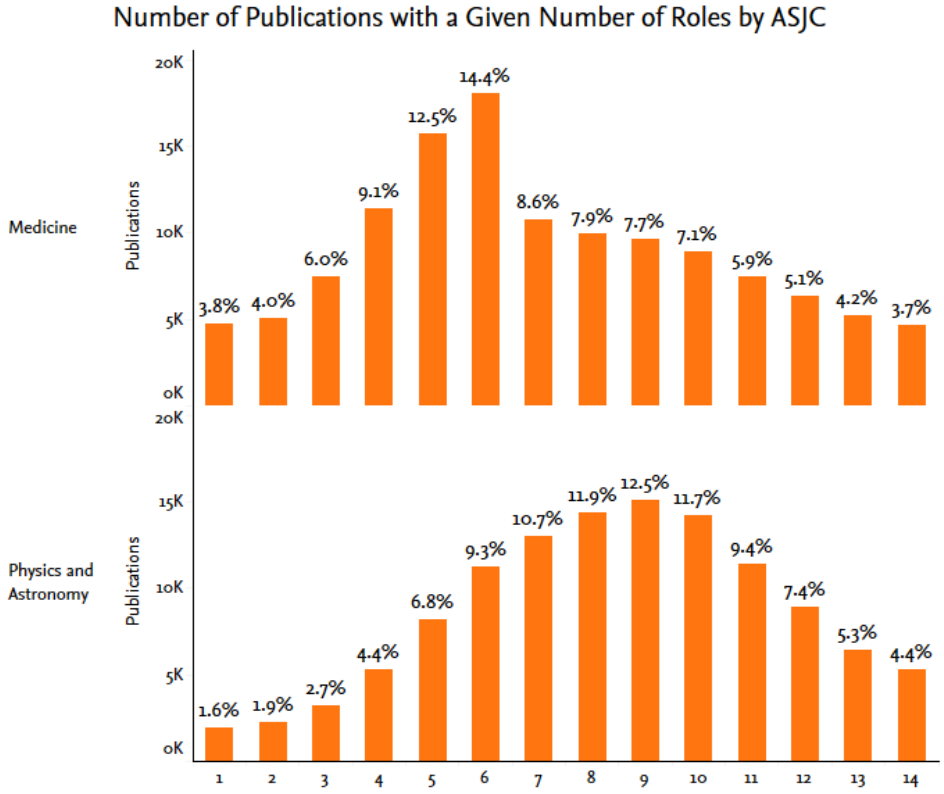
Therefore, publications with few and many roles are not common, meeting the expectations. Publications with very few roles should be naturally rare due to research usually requiring at least several of CRediT contributions.



Number of roles in publications

There are differences in the distributions of publications by number of roles across different ASJC fields.

For example, publications in Medicine tend to favour fewer number of roles than publications in Physics and Astronomy.

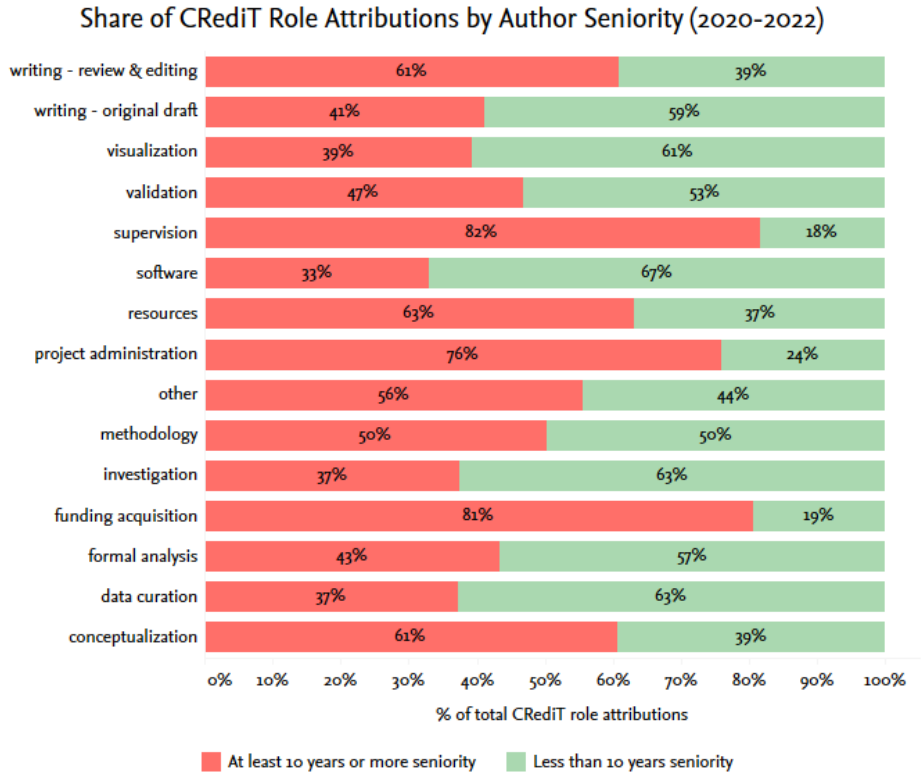


CRediT roles by author seniority

There are differences in the number of specific role attributions between senior authors with at least 10 years of publishing history and authors with less than 10 years of publishing history.

The data shows that senior authors tend to get acknowledged more for roles such as funding acquisition, supervision and project administration. In turn, less senior authors get acknowledged more for roles such as software, investigation, data curation and visualisation.

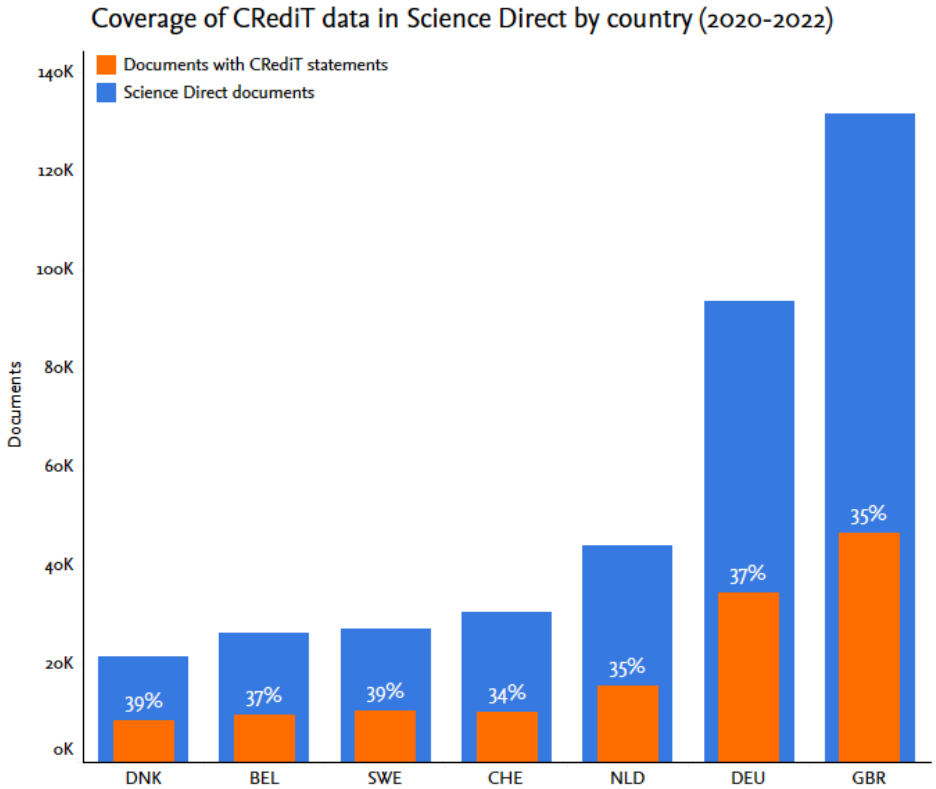
These observations corroborate that the dataset provides a realistic representation of role dynamics.



Coverage of CRediT data by country

The coverage of CRediT data in Science Direct varies across different countries. In 2020-2022, the Netherlands had 35% of publications in Science Direct with CRediT statements.

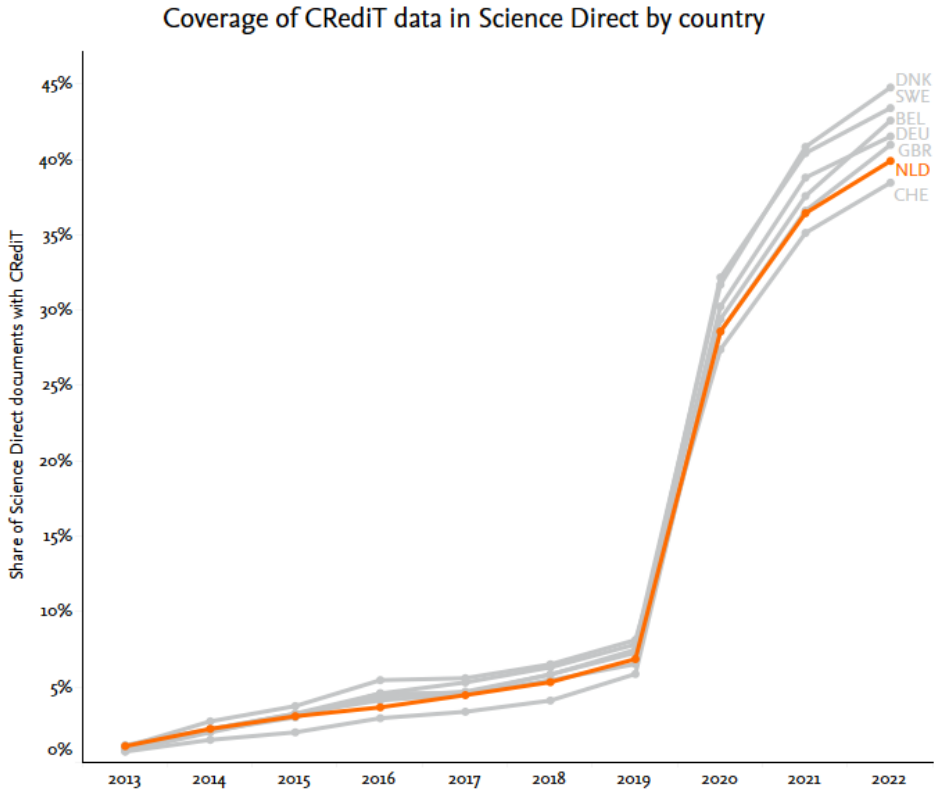
Denmark and Sweden have the highest coverage at 39%, while Switzerland has the lowest coverage among the compared countries at 33%.



Coverage of CRediT data by country

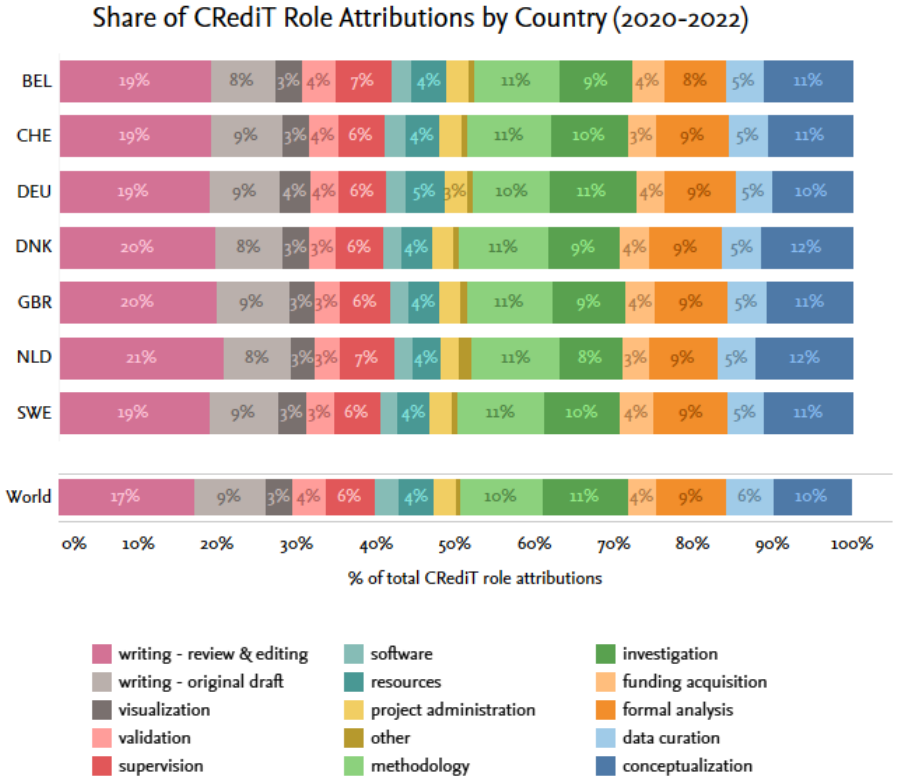
Analogously to the worldwide increase in CRediT coverage, the coverage data by country shows a similar pattern.

As of 2022, the Netherlands had 40% of Science Direct publications with CRediT statements. For Denmark, this figure is 45%.



Role attributions by country

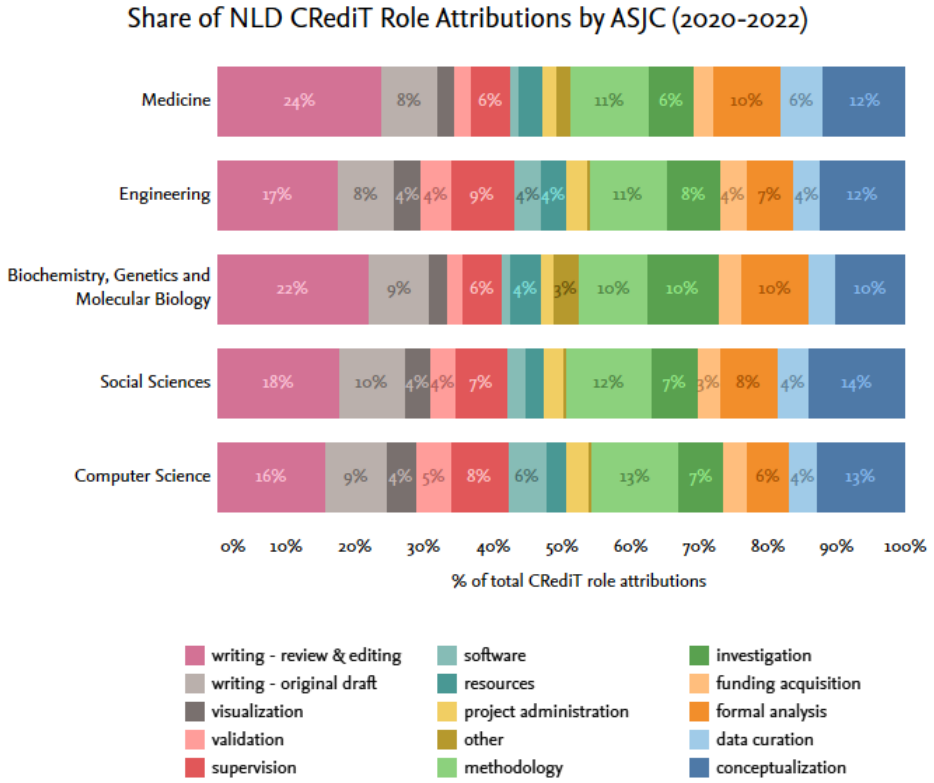
The share of specific CRediT role attributions is relatively consistent across countries and exhibits only slight variations when compared with the entire dataset (World).



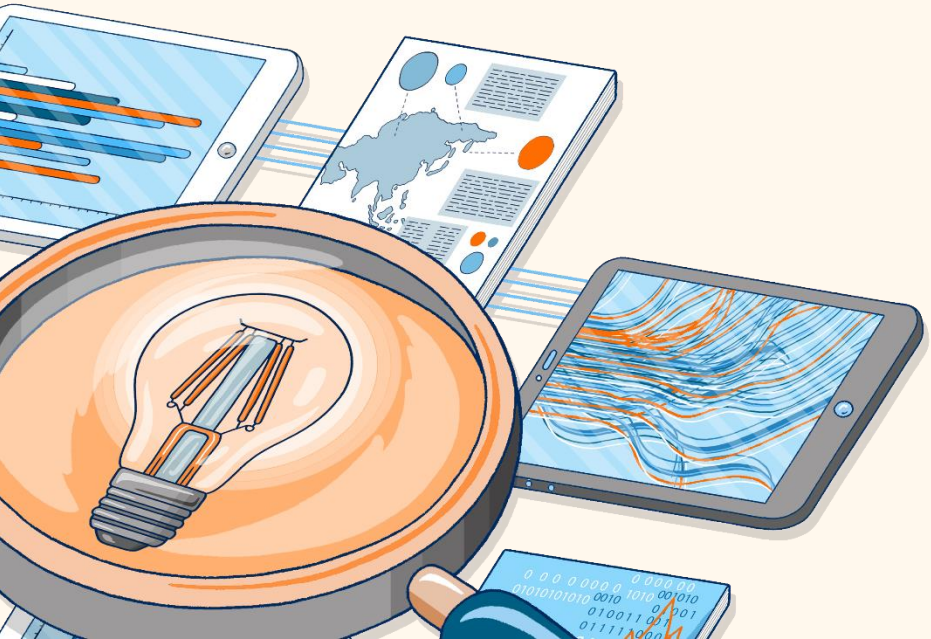
Role attributions in NLD by ASJC

There are, however, more notable variations in the share of specific CRediT roles depending on the ASJC subject field.

These variations are meaningful and align with some expectations about the relative importance of certain roles in different fields. For example, the share of the “software” role is the largest in Computer Science.



Conclusion



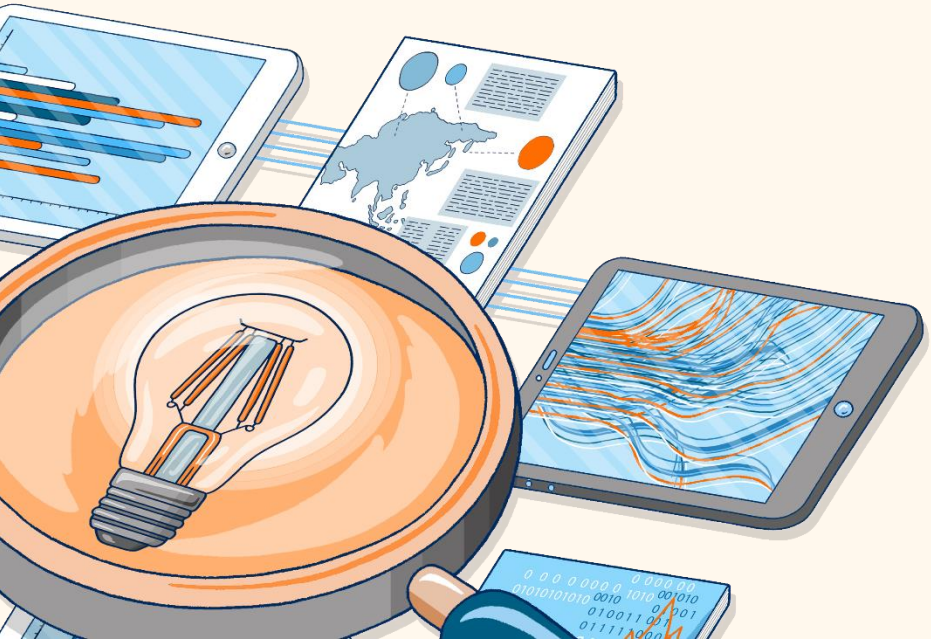
Summary of the findings

- The coverage of CRediT data in Science Direct is experiencing a strong uplift starting from 2020 and has reached nearly 50% of Science Direct documents in 2022.
- The coverage of CRediT data in the Netherlands is following a similar trend with 40% of its Science Direct content containing CRediT statements in 2022.
- The coverage varies across subject fields with the Multidisciplinary category having the highest coverage at 70%, which is likely due to this category housing many flagship journals with hight CRediT adoption.
- The distribution of publications with a given number of unique roles (1 to 14) is somewhat bell-shaped suggesting that publications with very few or very many roles are less common, although variations across different disciplines are possible.
- The analysis shows that the CRediT data captures meaningful aspects of author contribution dynamics. This is corroborated by the observed differences in the share of specific role contributions from junior and senior authors as well as differences across subject fields.

Opportunities for further analysis

- The analysis hints at multiple opportunities for further exploration. Investigating author contribution dynamics within smaller groups of researchers (e.g. departments) could be a window into understanding the unique characteristics of research teams, identifying their strengths and weaknesses.
- Further investigation of role attribution by gender and seniority can shed light on potential disparities within the academia. Such an inquiry may reveal patterns of inequality and inform strategies to foster a more equitable environment.
- Enhancing the understanding of author profiles based on their history of CRediT contributions can yield valuable insights. This can also involve analysing the permutations of role combinations authors undertake over time to understand how researchers develop in their careers.

Annex



The taxonomy: roles

1. Conceptualization: Ideas; formulation or evolution of overarching research goals and aims.

2. Methodology: Development or design of methodology; creation of models.

3. Software: Programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code components.

4. Validation: Verification, whether as a part of the activity or separate, of the overall replication/reproducibility of results/experiments and other research outputs.

5. Formal Analysis: Application of statistical, mathematical, computational, or other formal techniques to analyse or synthesize study data.

6. Investigation: Conducting a research and investigation process, specifically performing the experiments, or data/evidence collection.

7. Resources: Provision of study materials, reagents, materials, patients, laboratory samples, animals, instrumentation, computing resources, or other analysis tools.

The taxonomy: roles

8. Data Curation: Management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where it is necessary for interpreting the data itself) for initial use and later re-use.

9. Writing – Original Draft: Preparation, creation and/or presentation of the published work, specifically writing the initial draft (including substantive translation).

10. Writing – Review & Editing: Preparation, creation and/or presentation of the published work by those from the original research group, specifically critical review, commentary or revision – including pre- or post-publication stages.

11. Supervision: Oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team.

12. Project Administration: Management and coordination responsibility for the research activity planning and execution.

13. Funding Acquisition: Acquisition of the financial support for the project leading to this publication.

14. Visualization: Preparation, creation and/or presentation of the published work, specifically visualization/data presentation.