



Advantages of a data repository for publishing supplementary data in research articles

Poster abstract

Thomas Th. Kieselbach, Kristoffer Lindell, Olivia Ekman, Johanna Österåker and Carolin A. Rebernic

Umeå University Library, Scholarly Communication, 901 74 Umeå, Sweden

Contact: thomas.kieselbach@umu.se

This poster is accessible at the SciLifeLab Data Center at <https://doi.org/10.17044/scilifelab.23985792> and it has a CCBY 4.0 license.

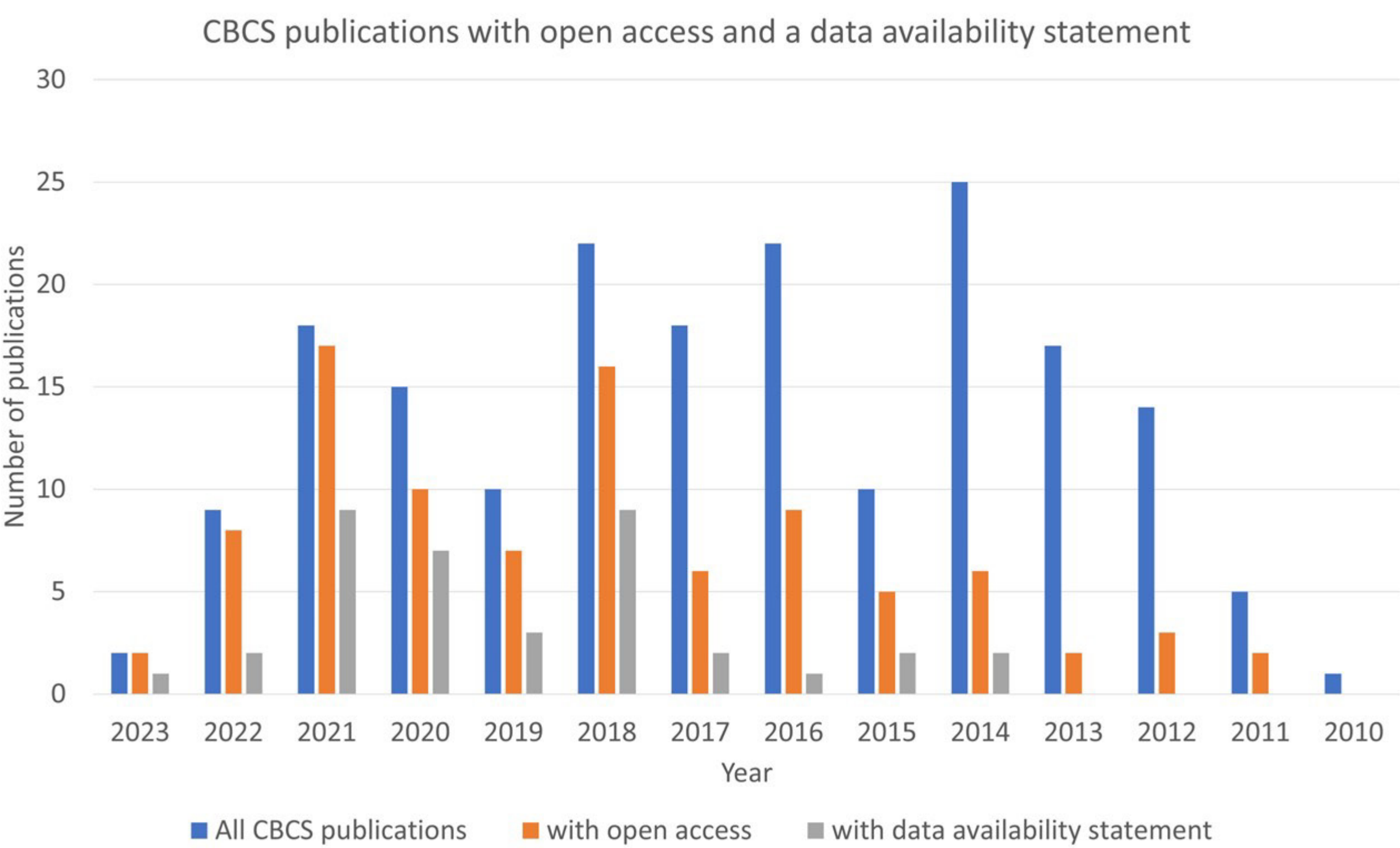


Contact Umeå University Library

In this study, we present an analysis of the publications listed on the homepage of the Swedish Chemical Biology Consortium (CBCS). We looked at some markers for open science including publication with open access, the presence of data availability statements and the application of the FAIR-guiding principles [1, 2]. Most recent CBCS publications are open access articles and many of them include a data availability statement. However, researcher do not seem to take advantage of the new opportunities that data repositories offer them for preparing research data for publications. It seems to be well established in the community to submit sequence data to Genbank and EMBL and structure models to the protein Data Bank and to the Cambridge Structure Data Bank. However, a large amount of research data is provided as supplementary material that frequently has no persistent identifiers and no metadata description [3]. We suggest considering data repositories to prepare supplementary data for publication in research articles because that provides many advantages. For instance, supplementary data can get a persistent identifier and become a citable research output [4]. Examples for suitable research data repositories are Figshare at SciLifeLab, Zenodo and the catalogue of the Swedish National Data Service.

References
[1] FORCE11 2017, The FAIR Data Principles, viewed 7 July 2023, <https://www.force11.org/group/force11/fairprinciples>.
[2] Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Bolten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouvman, J., Brookes, A. J., Clark, T., Crosas, M., Dilloo, J., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>
[3] Kieselbach, Thomas Thressa (2023). Analysis of CBCS publications for Open Access, data availability statements and persistent identifiers for supplementary data. SciLifeLab. Dataset. <https://doi.org/10.17044/scilifelab.23641749>
[4] Splawa-Neyman, Patrick (2018). Figshare and the FAIR data principles. Figshare. Journal contribution. <https://doi.org/10.6084/m9.figshare.7476428.v1>

Figure 1: Most recent CBCS publications are open access publications and many have a data availability statement.



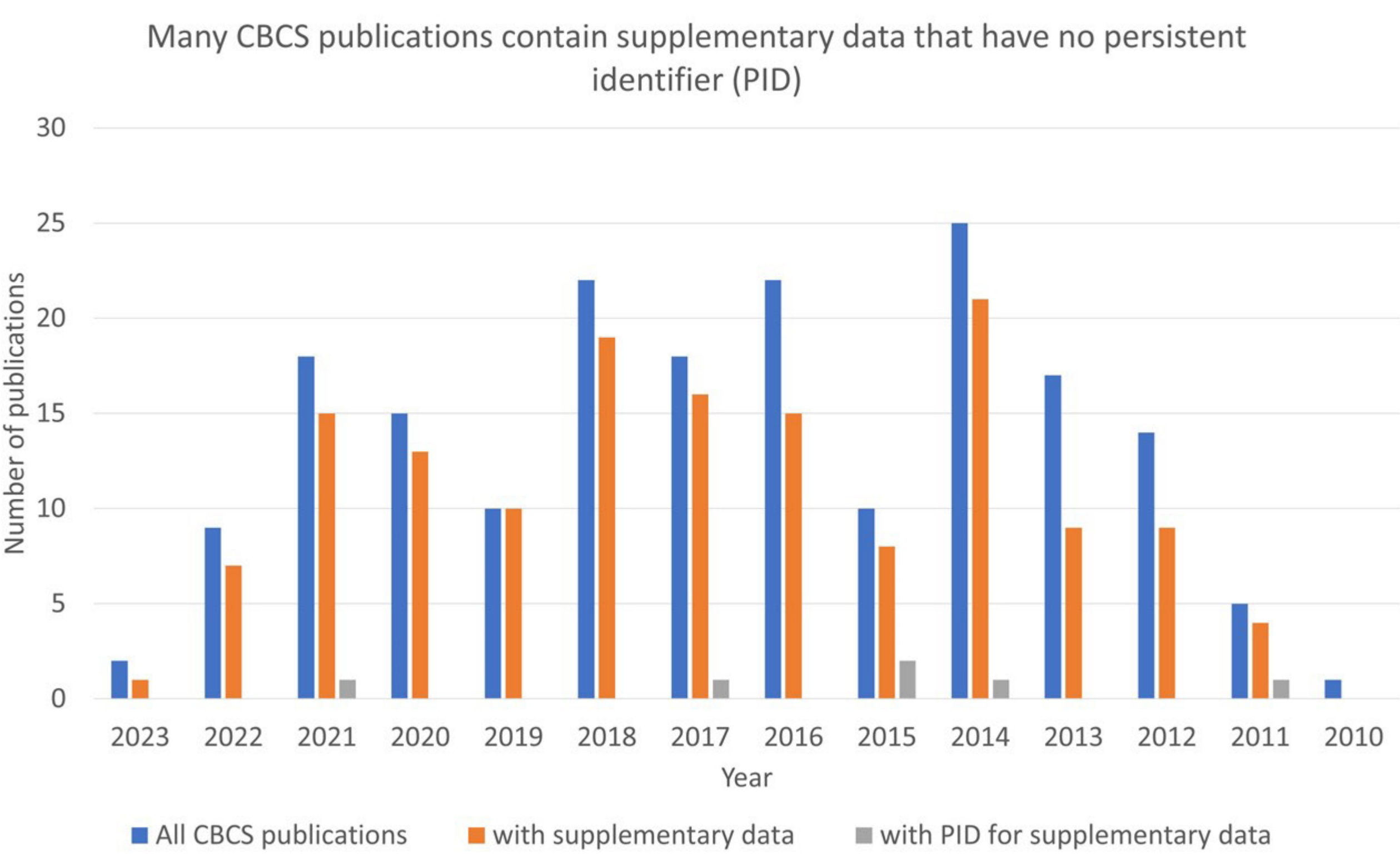
Advantages of a data repository for giving access to research data

Researchers use established databases such as Genbank, the Protein Data Bank and the Cambridge Structure Data Bank to share sequence and structure data. Other data that are relevant for scholarly publications are frequently provided in the form of figures and tables in the body text and as supplementary material.

That has disadvantages because supplementary material has in many cases no persistent identifier, no good metadata profile and is not machine readable. Using a trustworthy general repository such as Figshare at SciLifeLab, Zenodo or the catalogue of the Swedish National Data Service is a better solution to prepare data for publication.

- Here are some advantages of using a trustworthy repository for supplementary data.
1. You keep control of your data and who can access to them.
 2. You can assign a persistent identifier to your data.
 3. You can create a detailed metadata profile for your data.
 4. You can apply the FAIR guiding principles.
 5. You can maintain your data and create new versions of them.
 6. You can avoid broken links and make sure that your data are accessible.
 7. You can make your data be a citeable research output.

Figure 2: Many CBCS publications have supplementary data without a PID.



FAIR-principles in Figshare [4]

Principle	What Figshare offers to comply with a FAIR principle
F1	The description of every Figshare item includes a DOI on its landing page.
F2	Available fields for metadata are title, authors, description, keywords, categories, licence, item type, external references, funding, publisher and contact.
F3	Figshare is indexed by Google.
F4	The metadata include a DOI that points to the landing page of an item.
A1	Figshare uses HTTPS as a standardized protocol.
A1.1	HTTPS is an open and free protocol. Researchers can use private links when needed.
A1.2	Figshare supports Single Sign-on services, Security Assertion Markup Language and federations.
A2	Figshare allows people to create records that contain only metadata.
I1	Figshare provides different languages for communication and information exchange between different computer systems.
I2	Figshare uses the Australian and New Zealand Standard Research Classification Fields of Research (FoR) codes and controlled vocabulary for item types.
I3	Figshare allows creating collections of items and of their metadata.
R1	Researchers can add detailed descriptions of their data in the description field and in readme files.
R1.1	Figshare allows researchers to assign different licences to their data.
R1.2	For public items the creator is the citable author. Other authors can be added to the description.
R1.3	Mandatory metadata are DOI, authors, title, publisher, publication year and type of item.

Table 1. Properties of CBCS publications*

Year	All CBCS publications	With open access	With data availability statement	With supplementary data	With PID for supplementary data
2023	2	2	1	1	0
2022	9	8	2	7	0
2021	18	17	9	15	1
2020	15	10	7	13	0
2019	10	7	3	10	0
2018	22	16	9	19	0
2017	18	6	2	16	1
2016	22	9	1	15	0
2015	10	5	2	8	2
2014	25	6	2	21	1
2013	17	2	0	9	0
2012	14	3	0	9	0
2011	5	2	0	4	1
2010	1	0	0	0	0

* Data describe the status of 6 July 2023.

FAIR guiding principles by Force11 [1, 2]

- To be Findable:**
- F1. (meta)data are assigned a globally unique and eternally persistent identifier.
 - F2. data are described with rich metadata.
 - F3. (meta)data are registered or indexed in a searchable resource.
 - F4. metadata specify the data identifier.

- To be Accessible:**
- A1 (meta)data are retrievable by their identifier using a standardized communications protocol.
 - A1.1 the protocol is open, free, and universally implementable.
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
 - A2 metadata are accessible, even when the data are no longer available.

- To be Interoperable:**
- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
 - I2. (meta)data use vocabularies that follow FAIR principles.
 - I3. (meta)data include qualified references to other (meta)data.

- To be Re-usable:**
- R1. (meta)data have a plurality of accurate and relevant attributes.
 - R1.1. (meta)data are released with a clear and accessible data usage license.
 - R1.2. (meta)data are associated with their provenance.
 - R1.3. (meta)data meet domain-relevant community standards.