

Forskningsdatapubliceringar vid Stockholms universitet – var finns de?


Maria Almbro | Lisa Olsson

Forskningsdatateamet, Stockholms Universitet

Searching For Research Data

An Assessment of Data Publication Practises at Stockholm University

Maria Almbro & Lisa Olsson
maria.almbro@su.se lisa.olsson@su.se



Background

Researchers are increasingly encouraged and required by funders, journals, universities and governments to make research data FAIR¹ and openly available, preferably in a data repository. In Sweden, the aim is an open scientific system by 2026. However, little is currently known about whether and how researchers make their research data available.

Methods

We randomly sampled **10% of all research articles** across all research fields published in 2022 from Stockholm University.

Manually assessed 364 research articles

63 articles were excluded (e.g. review, theory, opinion papers)

301 empirical articles were included in the assessment of research data availability

Results

Question	Yes (%)	No (%)	Uncertain (%)
Does the article have a data availability statement?	45%	55%	0%
Is a DOI ² for the data provided in the article?	18%	82%	0%
Do articles that provide a DOI for the data actually contain data?	53%	20%	27%
Does the article have supplementary files?	46%	54%	0%
Do supplementary data files actually contain data?	23%	43%	34%
Does the article say data is available upon request?	19%	81%	0%
Was the reason(s) for restricting data availability reasonable e.g. sensitive data?	46%	54%	0%
Does the article include links to other data sources?	29%	71%	0%

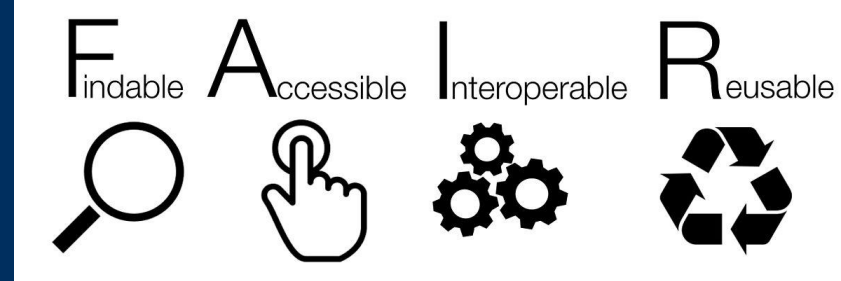
Conclusions

Researchers **do not** publish data underpinning research findings consistently. However, almost half of all sampled research articles provide a Data Availability Statement. Few publish data with a DOI in a repository, as a supplement in the article or link to other data sources. However, there is a large degree of uncertainty because of the difficulty of assessing what constitutes data and what is "as closed as is necessary" in various fields.

¹FAIR: Findable Accessible Interoperable Reusable
²DOI: Digital Object Identifier

Acknowledgements: Gabor Schubert, Stockholm University Library

REGERINGENS MÅLBILD



” ...vetenskapliga publikationer, som är ett resultat av forskning som finansieras med offentliga medel, ska vara omedelbart öppet tillgängliga med verkan från 2021. Vad gäller forskningsdata ska omställningen vara genomförd fullt ut senast 2026, vilket innebär att forskningsdata ska göras tillgängligt så öppet som möjligt och så begränsat som nödvändigt.

– Regeringens forskningsproposition 2020/21:60, s. 101

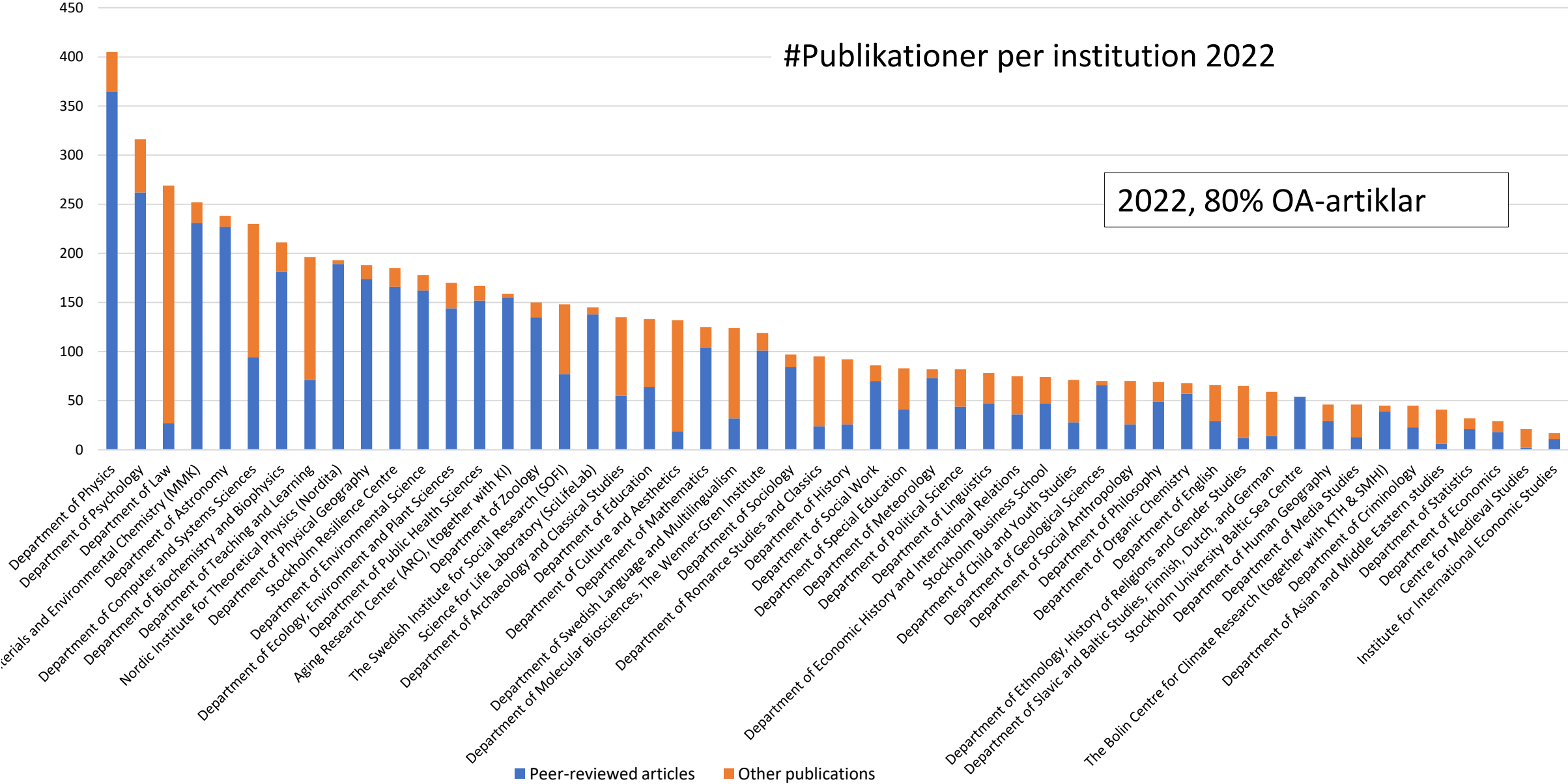
Stockholms universitet

- Ett av de största svenska lärosätena
 - 5,600 anställda
 - 29,200 studenter
 - 1,300 doktorander
- Open Science-profil
- 4 fakulteter, heterogent & decentraliserat universitet
 - Naturvetenskapliga området (Naturvetenskapliga fakulteten)
 - Humanvetenskapliga området (Humanistiska-, juridiska- & samhällsvetenskapliga fakulteten)
= 51 institutioner, centra och institut
- 3,700+ artiklar/år



#Publikationer per institution 2022

2022, 80% OA-artiklar



Acknowledgement: Gabor Schubert, Stockholms universitetsbibliotek

VAD ÄR STATUSEN FÖR ÖPPNA DATA VID STOCKHOLMS UNIVERSITET?

- Mycket oklart!
- Olika praxis inom olika discipliner
- Svårt att mäta enhetligt
- Kvantitativa ansatser har sina brister
- Data: ~5-10% (uppmätt via DOIs) – kan vara mer eller mindre



FÖRLAGENS "DATA AVAILABILITY" POLICIES

- Springer Nature
 - Uppmuntrar tillgängliggörande utan dröjsmål
- Taylor & Francis
 - Uppmuntrar tillgängliggörande
- Elsevier
 - Uppmuntrar tillgängliggörande
- Wiley
 - Uppmuntrar/förväntar/kräver tillgängliggörande
- SAGE
 - Rekommenderar tillgängliggörande

	Data availability statement is published ¹	Data has been shared ²	Data has been peer reviewed ³	Example Wiley journals
Encourages Data Sharing	Optional	Optional	Optional	
Expects Data Sharing	Required	Optional	Optional	British Journal of Social Psychology
Mandates Data Sharing	Required	Required	Optional	Ecology and Evolution
Mandates Data Sharing and Peer Reviews Data	Required	Required	Required	Geoscience Data Journal American Journal of Political Science

¹ A data availability statement confirms the presence or absence of shared data.

² Links to data in data availability statements are checked to ensure they link to the data that the authors intended. If data have been shared in a data repository, the data availability statement includes a permanent link to the data. Shared data is also cited.

³ Quality and/or replicability of linked data are peer reviewed. Depending on the journal, this may be to peer review the quality of the data by ensuring that the results in the paper and the data in the repository align (for example, sample sizes and variables match), or it may be to peer review the replicability of the data to ensure that the claims presented in the journal article are valid and can be reproduced.

Example: Wiley's Data Sharing Policies (<https://authorservices.wiley.com/author-resources/Journal-Authors/open-access/data-sharing-citation/data-sharing-policy.html>)



VAD VI REKOMMENDERAR

- Publicera data som underbygger din artikel i ett repository
- Lägg datapublikationens DOI i ett Data Availability Statement i din artikel
- Nämn artikelns DOI i din datapublikation i repositoryet
- Känsliga data? Publicera metadata – inga datafiler laddas upp



Data and Code Availability

All code and statistical software commands are available online ([https://github.com/LNDG/Mansson_etal_2022_BiolPsychiatry ↗](https://github.com/LNDG/Mansson_etal_2022_BiolPsychiatry)). Owing to ethics constraints, we cannot at present make the raw patient data openly available; please contact the first author (KNTM) to discuss potential routes to data access.

VÅR UNDERSÖKNING

- Valde slumpmässigt ut 10% av alla sakkunniggranskade artiklar från Stockholms universitet 2022 (DiVA)
- Bedömde manuellt huruvida empiriska data underbyggde artiklarna – förväntades data finnas?
- Frågeställningar:
 - Finns Data Availability Statement?
 - Hur har data tillgängliggjorts?
 - Publiceras och länkas data med DOI från artikeln, som vi rekommenderar?
 - Finns supplementary material som pdf/doc bifogade till artikeln? (Innehåller dessa data?)
 - Innehåller artikeln länk till existerande datakällor?
 - Är data "available upon request"? (Är data känslig eller ej?)

RESULTAT

- 370 artiklar, 6 otillgängliga → 364 artiklar
- 301 (81%) artiklar bedömdes vara empiriska (baserade på data)
- Av 301 artiklar
 - 45% hade ett Data Availability Statement (eller liknande fras)
 - 21% refererade inte till underliggande data på något vis
 - 19% uppgav att data var "available upon request" (varav ca 50% uppskattades vara känsligt material som inte kan delas)

TABLE OF CONTENTS

Abstract

Introduction

Materials and methods

Results

Discussion

Data availability statement

Ethics statement

Author contributions

Funding

Acknowledgments

Conflict of interest

Publisher's note

Supplementary material

References

RESULTAT (2)

Vad gäller tillgängliggörande av data:

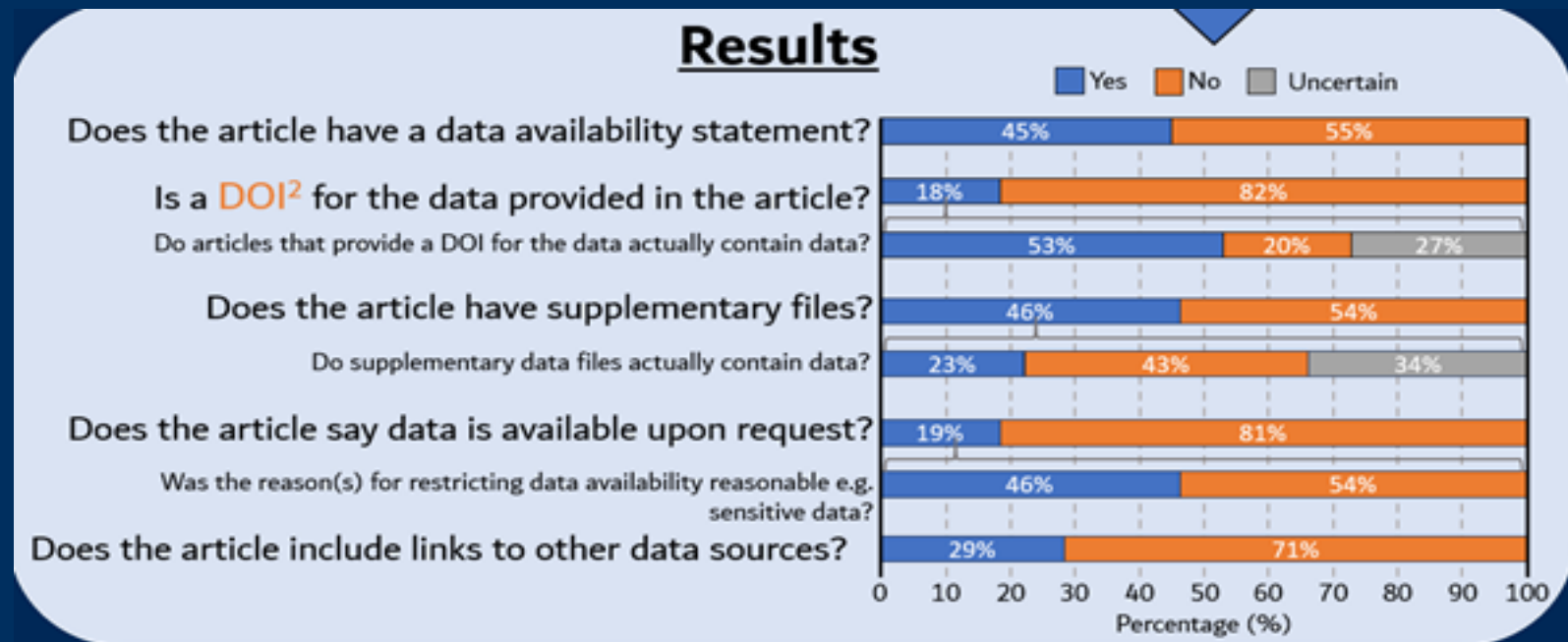
- författare bifogade ofta "supplementary material" till artiklar (46% av 301). Sällan data!
→ 11% av 301 artiklar hade data i supplementary material
- författare länkade ganska ofta till en existerande datakälla från sin artikel (29% av 301) Ofta data (inte alltid direkt tillgänglig)
- författare publicerade något med en DOI (18% av 301) och länkade till det från sin artikel

Innehöll data i cirka hälften av fallen

→ 10% av de 301 artiklarna hade data persistent länkade till dem med en DOI

RESULTAT (3)

- Totalt hade 39% av de 301 artiklarna åtminstone någon referens till underbyggande data (publicerad med DOI, tillgängligt i supplementary material eller via länk till en annan datakälla)
- Betyder inte att 39 % av artiklarnas data tillgängliggjorts så ”*öppet som möjligt och så begränsat som är nödvändigt*”



SLUTSATSER

- Öppna Data <39% – Stockholms universitet är långt ifrån målet om Öppna och FAIR Data
- Forskare vid SU publicerar inte sin data konsekvent – ”data” betyder olika saker i olika discipliner
- Data Availability Statements är vanligt förekommande (45%)
- Få (10 %) publicerar data med DOI i repositorer, som vi rekommenderar
- Fler (29%) publicerar och länkar till data och/eller kod (utan DOI) i databaser eller länkar till data i existerande källor (ex. registerdata)
- Supplementary material är vanligt (46%), men innehåller sällan data och underhålls ej
- Vår manuella och kvalitativa metod ”fångar” mer data än kvantitativa ansatser som kopplar artiklar till DOI enbart (39% vs 5-10%)

HUR KAN VI FÖRBÄTTRA PRAKTIKER?

- Forskare
 - Inkludera Data Availability Statements i artiklar och använd DOIs (eller likvärdigt) för att länka ihop publikationer
 - Skilj data från supplementary material och undvik att lägga data i bilagor. Beskriv och publicera data i repositorier
- Förlag/tidskrifter
 - Informera forskare om hur de kan tillgängliggöra data och extra material på bästa sätt
 - Uppmuntra forskare att kurera data väl och publicera det med DOIs
 - Underhåll/hantera DOIs för data som en del av artikelmetadata
- Lärosäten
 - Informera forskare om forskningsdatapublicering och FAIR
 - Erbjud datakureringstjänster och infrastruktur till forskare

Meriterande
och
uppmuntrat av
finansiärer

TACK!

