



Spårbara forskningsdata

Metoder och rekommendationer för ökad spårbarhet och synlighet för svenska datapubliceringar

André Jernung (SND), Stina Johansson (GU) och
Therese Tikkanen (Chalmers), 2025-10-22



SND

Svensk nationell datatjänst | Göteborgs universitet - Chalmers tekniska högskola - Karolinska Institutet - Kungliga Tekniska högskolan - Lunds universitet - Stockholms universitet - Sveriges lantbruksuniversitet - Umeå universitet - Uppsala universitet

Projektets struktur

Chalmers

Therese Tikkanen
Urban Andersson

SND

André Jernung
Olof Olsson

KTH

Lina Andrén
Mattias Vesterlund

GU

Stina Johansson



Bakgrund och syfte

- Uppföljning
 - Överblick
 - Behov av rekommendationer
-
- Nulägesbild
 - Brister och behov
 - Förbättrad uppföljning, utvärdering och kvalitetsgranskning

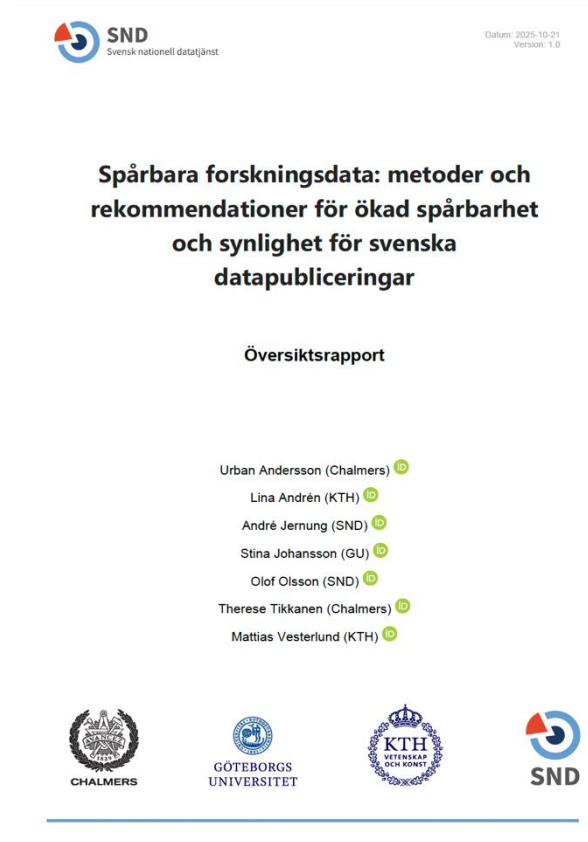


Resultat och leveranser

- Översiktsrapport
- Gapanalys
- Rekommendationer bevakning och registrering
- Rekommendationer till forskare
- Tekniska lösningar
- Workshop

Översiktsrapport

- Översikt:
 - Nationella krav
 - Tekniska förutsättningar
 - Stödorganisation
 - Tillgängliggörande / publicering
- Tidigare litteratur kring datapublicering
- Kartläggning / enkät



Andersson, U., Andrén, L. J., Jernung, A., Johansson, S., Olsson, O., Tikkanen, T., & Vesterlund, M. (2025). Spårbara forskningsdata: metoder och rekommendationer för ökad spårbarhet och synlighet för svenska datapubliceringar - Översiktsrapport (1.0). Swedish National Data Service. <https://doi.org/10.5281/zenodo.17365330>

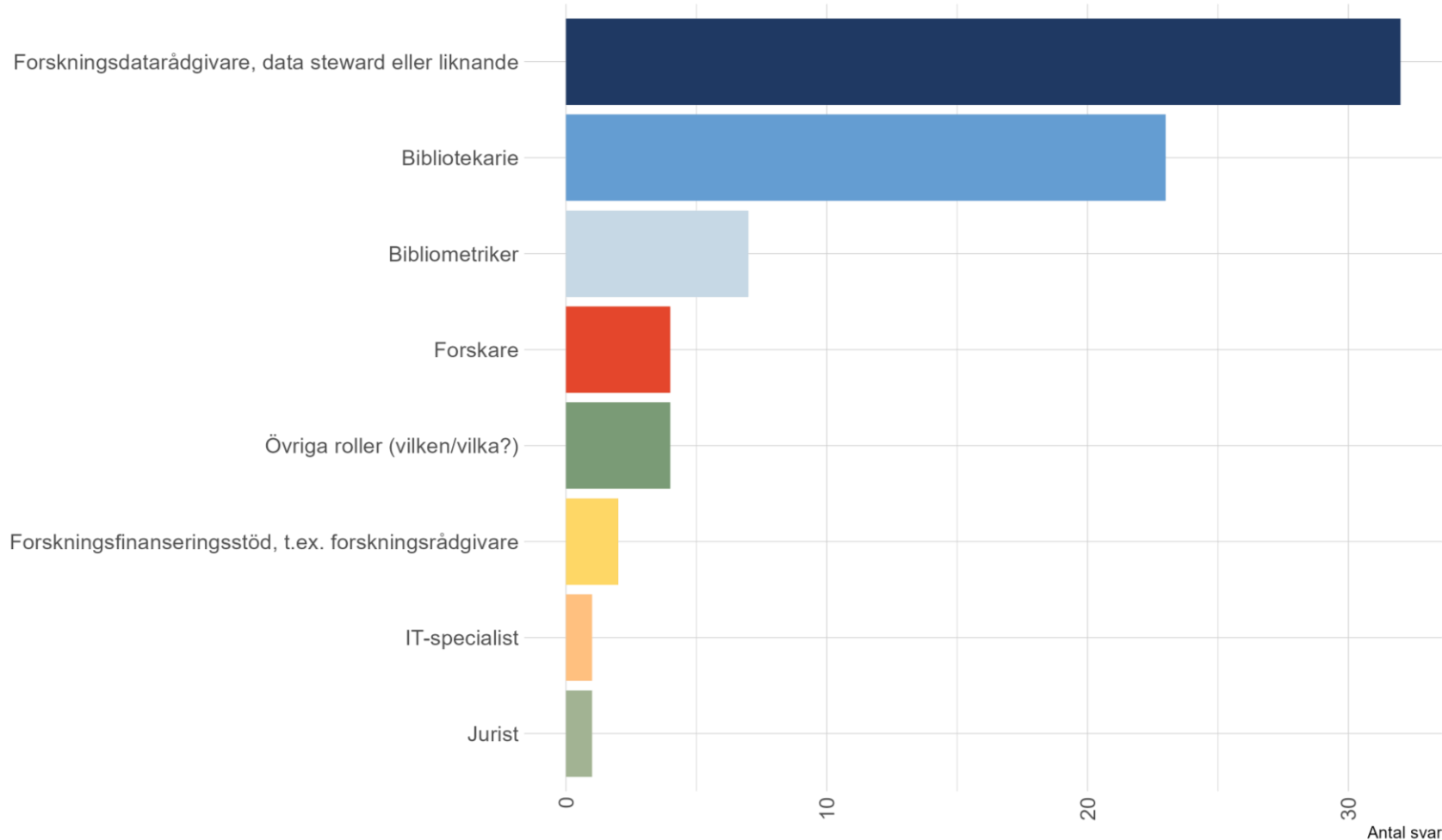


bit.ly/sparbaradata

Enkät: Att publicera och hitta forskningsdata som skapats vid svenska lärosäten och forskningsorganisationer

Roller, antal svar per roll

Flera svar per person var möjliga



- 54 respondenter
- Stor andel fritextsvar
- Analys genom kvalitativ kodning (Taguette)



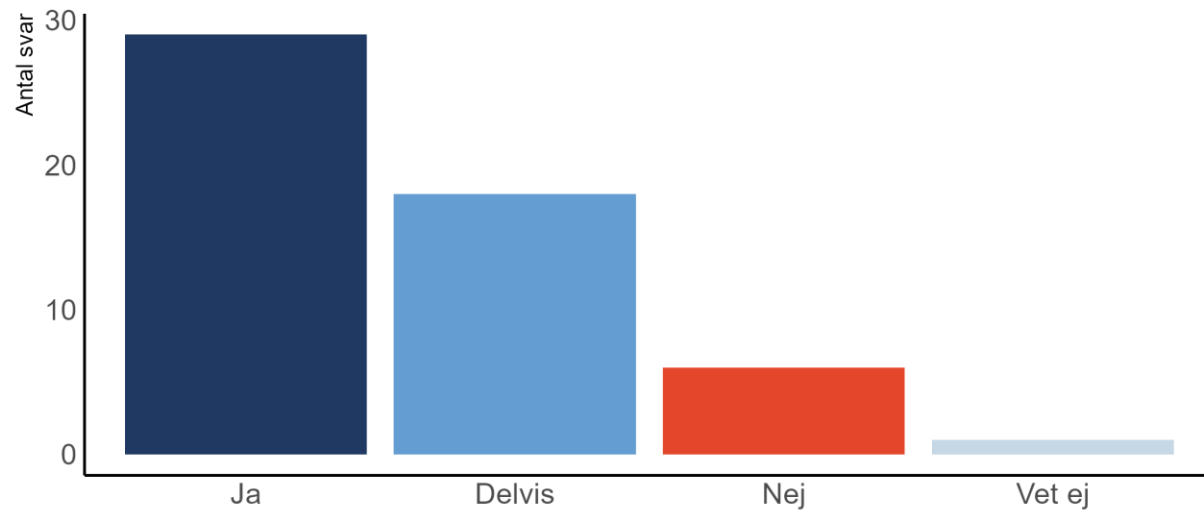
Tema: Infrastruktur och processer

- Att publicera data upplevs som komplext och svårt
- Datapublicering planeras ofta för sent i processen
- Många organisationer saknar system (CRIS-system eller publikationsdatabaser) där datapubliceringar kan registreras
- Det är svårt att publicera data med begränsad tillgång, antingen på grund av att rutiner inte är på plats eller för att det kräver för mycket resurser
- Frågan kring långsiktig lagring och arkivering av data inte löst/under arbete
- Publicerade data går inte alltid att citera, och de citeras eller hänvisas inte heller på ett enhetligt sätt

”Vi har just nu Data Monitor från Elsevier som kan hitta en del dataset, men den läggs ner i juli. Någon motsvarighet till den tjänsten skulle behövas.[...]”

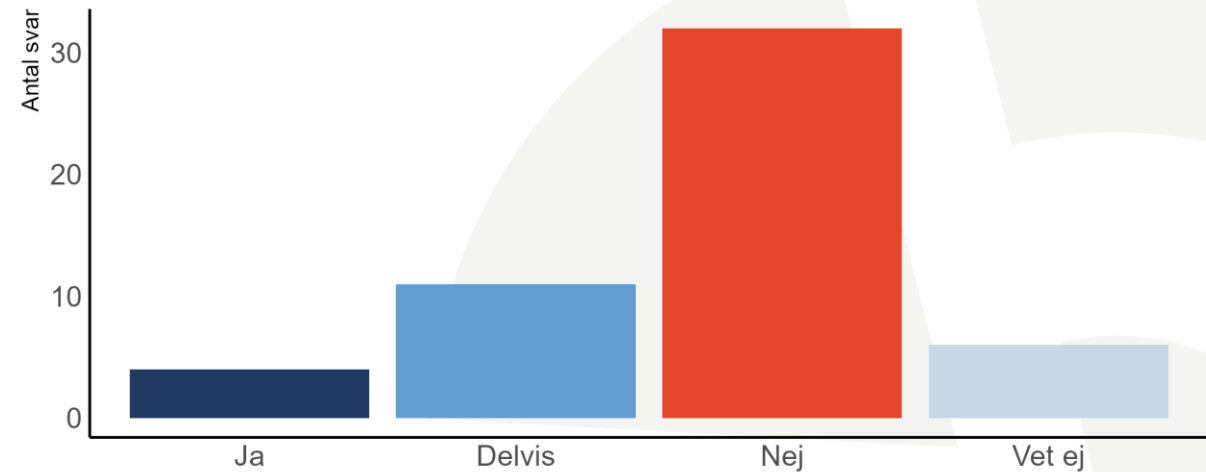
Kurering och Granskning

Arbetar din organisation med aktiv kurering eller granskning av datapubliceringar och tillhörande metadata?



Registrering

Gör ni någon registrering eller katalogisering av datapubliceringar som kommer till er kännedom, t.ex. i ett CRIS-system, institutionsrepositorium, publikationsdatabas eller annat liknande lokalt system?

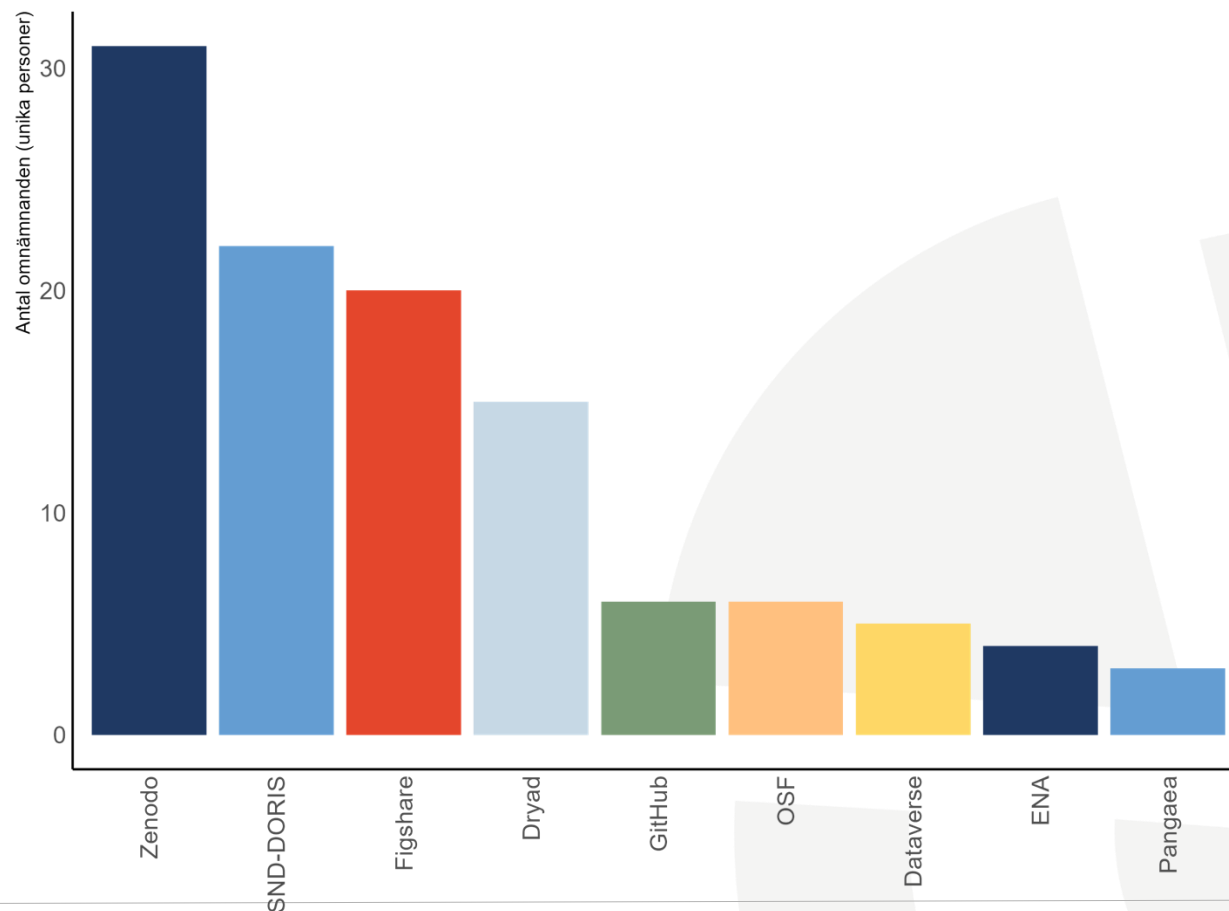


- Flera organisationer kurerar datapubliceringar, men det görs i några få repositorer/infrastrukturer

”Allra svårast tror jag är att hitta sätt att försäkra sig om att man hittat data som publiceras i mindre, ämnesspecifika repositorer.”

Kända Publiceringsplattformar

Vilka repositorer eller andra publiceringsplattformar för forskningsdata som era forskare publicerar sig i känner ni till?



Tema: Standardisering och interoperabilitet

- Metadata är bristfälliga
 - Ofullständig eller tvetydig affilieringsinformation
 - Minimikrav varierar – viktig information kan saknas
 - Bristande användning av beständiga identifierare även när sådana existerar och stöds
- Metadata är ostandardiserade
 - Avsaknad av relationer till andra objekt (t.ex. datapublicering – artikel – projekt)
 - Avsaknad av viktig information som ämne eller licens
 - Olika tolkningar av standarder gör att de implementeras olika
- Data availability statements
 - Dessa är inte standardiserade och deras användbarhet och sanningsenlighet ifrågasätts



Tema: Nationell samordning

- Många efterfrågar ett nationellt system för att samla in information om forskningsprojekt (inkl. datapubliceringar)
- *"Svenska [organisationer] och forskningsfinansiärer är inte bra på att dela information om forskningsdata från projekt vid svenska lärosäten med världen genom t.ex. OpenAIRE Graph."*
- Man ser även ett behov av ett nationellt grepp i frågan om incitament och meritering



Tema: Ledning och resurser

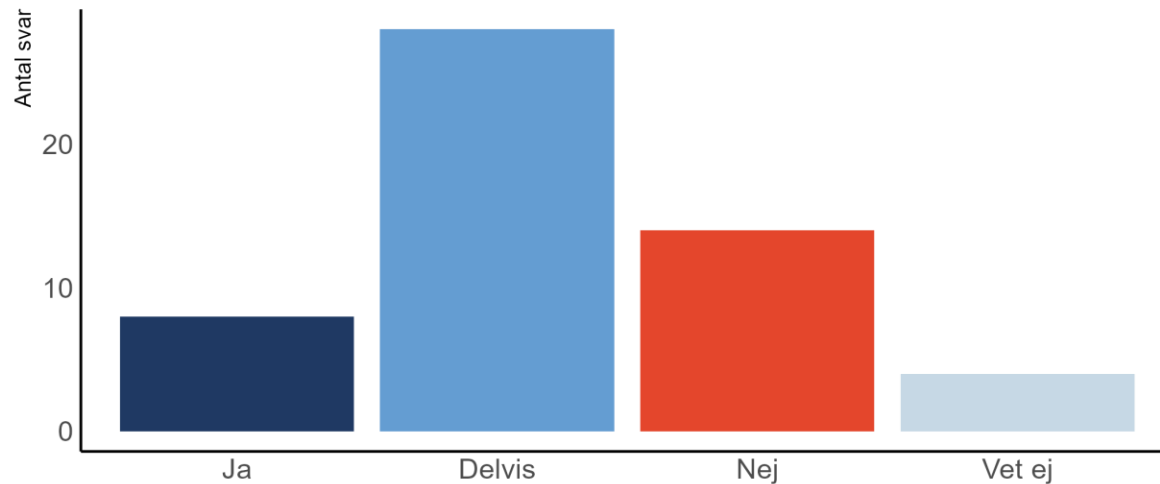
- Få har fått ett tydligt uppdrag om att aktivt registrera och rapportera datapubliceringar (trots öppna datalagen)
- Brist på tid, resurser, kompetens, både i stödorganisationer och bland forskare
- Datapublicering kan vara en lågt prioriterad fråga för forskaren

Tema: Uppföljning och incitament

”Jag upplever det som att vi arbetar väldigt mycket i motvind. Samma fråga ställs hela tiden ’varför ska vi göra det här’ ’vad är meningen’[...]”

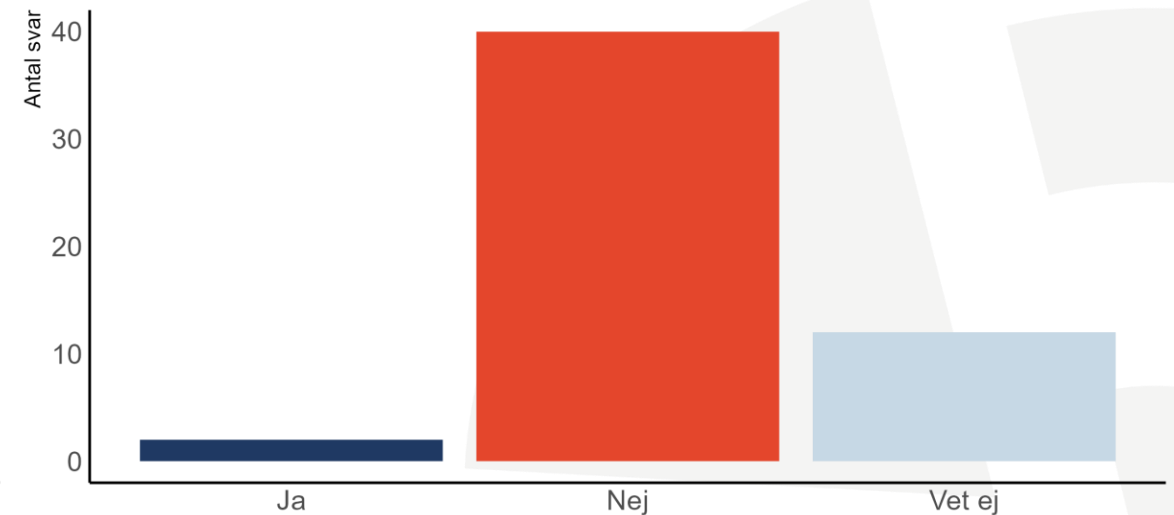
Incitament

Upplever du att det finns incitament för att publicera forskningsdata (oavsett tillgänglighetsnivå)?



Indikator

Har din organisation tagit fram någon form av indikator/KPI för forskningsdata?





Tema: Vägen mot en öppen vetenskapskultur

- Långsiktig kulturförändring; framväxandet av nya praktiker tar tid
- Citeringsmetoder för data måste uppdateras (tidskrifter och forskare) - vem skulle inte vilja se datapubliceringar i referenslistan?
- Initiativ runt öppen forskningsinformation (t.ex. Barcelona Declaration on Open Research Information) kan avhjälpa vissa problem
- Uppföljning, incitament och nya meriteringsprocesser;
- Men hur? Olika åsikter om kvantitativa indikatorer vs kvalitativ utvärdering

Krav
Policy



5. Gör det obligatoriskt

Utvärdering
Erkännande



4. Gör det meriterande

Gemenskaper
Föregångspersoner



3. Gör det till förebild

Användarvänlighet
Vägledning



2. Gör det enkelt

Infrastruktur
Processer



1. Gör det möjligt



Strategier för utveckling av metodik

- Breda målgruppers behov
 - Behov hos organisationer med små resurser
 - Hitta det som är ordnat innan vi släpper in kaos
 - Långsiktig ambition om metodstandardisering
-
- Enkla grundverktyg
 - Anpassningsbart – modularitet
 - Verktygskedja – flerpassmetodik
 - Tillgänglighet – Möjliggöra olika användargränssnitt

Vilka stora metadatakällor finns?

- Via API:er
 - DataCite
 - Crossref
 - OpenAIRE
 - OpenAlex
 - Scholix/ScholXplorer
- Via metadatadumpar
 - ORCID
 - Open Global Data Citation Corpus
 - DataCite
 - Crossref
 - m.fl.



Rent praktiskt



- DOI-metadata från DataCite som stomme
 - Bygga på med stökigare källor
 - Metadatakvalitet – hålla det värsta ute
 - Var varsam med exkludering av materialtyper
-
- I dagsläget två API:er som metadatakällor
 - DataCite
 - OpenAIRE Graph

roagg – Research output aggregator

- Första passet i verktygskedjan
- Skript som hittar datapubliceringar via affiliering
- Möjliggör bevakning via tillgängliga API:er
- Matchning på ROR ID och fritext – beroende av bra stränglistor
 - Creators/contributors, Publisher, ...
- Kod: <https://github.com/snd-sweden/research-output-aggregator>
- Datadumpar: <https://github.com/snd-sweden/research-output-aggregator-output/>

```
tests > name-lists > ⌵ chalmers.txt
1  chalmers universit*
2  chalmers hogskola*
3  chalmers university of technology*
4  universit* chalmers
5  university of chalmers
6  onsala *
7  chalmers onsala
8  chalmers tech*
9  chalmers tekn*
10 department * chalmers
```

SND Research Output Aggregator GRID

Simple tool for viewing and filtering output from the [SND Research Output Aggregator](#).

i This grid only shows a limited set of columns from the full output.
Download the full CSV for the selected organisation using the "Download" button.

Select organisation:

Uppsala University (uu)   Download uu.csv

	Doi	Client Name	Public...	Resource ...	Title	Publisher	Is Publisher	Is Latest Version	Is Concept Doi	Created At	Updated At
			☐				☐		☐		
☐	10.5281/zenodo.25333...	Zenodo	2019	Text	Feedback to cOAlition S on the Plan S Implementation Document	Zenodo	0	1	0	2019-01-07T20:57:51Z	2020-07-15T22:41:17Z
☐	10.5281/zenodo.25333...	Zenodo	2019	Text	Feedback to cOAlition S on the Plan S Implementation Document	Zenodo	0	0	1	2019-01-07T20:57:52Z	2020-07-15T22:41:16Z
☐	10.5061/dryad.19mn6...	DRYAD	2019	Dataset	Data from: Modelling durophagous predation and mortality rates from the fossil record of gastropods	Dryad	0	1	0	2019-01-07T23:04:46Z	2025-07-22T06:27:01Z
☐	10.5061/dryad.m66b10q	DRYAD	2019	Dataset	Data from: Climbing adaptations, locomotory disparity and ecological convergence in ancient stem ‘kangaroos’	Dryad	0	1	0	2019-01-11T18:14:45Z	2025-07-22T09:43:11Z
☐	10.5061/dryad.cm042k3	DRYAD	2019	Dataset	Data from: Effects of photoperiod on life-history and thermal stress resistance traits across populations of Dr...	Dryad	0	1	0	2019-01-11T19:40:41Z	2025-07-22T06:29:13Z
☐	10.5061/dryad.m073h5g	DRYAD	2019	Dataset	Data from: Birds with high lifetime reproductive success experience increased telomere loss	Dryad	0	1	0	2019-01-11T20:05:28Z	2025-07-22T00:25:31Z
☐	10.5281/zenodo.25381...	Zenodo	2019	Software	joshday/OnlineStats.jl: Better Extrema	Zenodo	0	1	0	2019-01-11T20:24:01Z	2024-09-10T07:36:36Z
☐	10.5281/zenodo.25388...	Zenodo	2019	Software	HITS-MBM/gromacs-developments: v2016.5 Sliced Pull	Zenodo	0	1	0	2019-01-14T02:14:03Z	2023-04-25T15:28:53Z
☐	10.5281/zenodo.25388...	Zenodo	2019	Software	HITS-MBM/gromacs-developments: v2016.5 Sliced Pull	Zenodo	0	0	1	2019-01-14T02:14:11Z	2023-04-25T15:28:54Z
☐	10.5061/dryad.4s5bf8d	DRYAD	2019	Dataset	Data from: The efficacy of good genes sexual selection under environmental change	Dryad	0	1	0	2019-01-16T18:53:12Z	2025-07-21T21:00:09Z
☐	10.17044/nbis/g000007	SciLifeLab Data Rep...	2019	Dataset	Data from: RNA Sequencing Provides Novel Insights Into the Transcriptome of Aldosterone Producing Adeno...	Uppsala University	1	1	0	2019-01-18T15:41:28Z	2020-07-17T21:49:33Z
☐	10.5878/8t28-pv80	Swedish National Da...	2025	Dataset	Land reflection seismic, hydrogeological and magnetic study of an area prone to quick-clay landslides in sout...	Uppsala University	1	0	1	2019-01-19T02:30:30Z	2025-09-22T10:21:28Z
☐	10.5878/acbv-h350	Swedish National Da...	2019	Dataset	Land reflection seismic, hydrogeological and magnetic study of an area prone to quick-clay landslides in sout...	Uppsala University	1	0	0	2019-01-19T02:30:32Z	2025-09-22T08:56:08Z
☐	10.17910/b7.828	New York University ...	2019	Dataset	the BasicChild Project	Databrary	0	1	0	2019-01-20T07:02:20Z	2021-12-26T07:08:13Z
☐	10.5281/zenodo.25467...	Zenodo	2019	Software	joshday/OnlineStats.jl: Bug fixes	Zenodo	0	1	0	2019-01-22T14:24:14Z	2024-09-10T07:36:32Z
☐	10.5061/dryad.fb17m59	DRYAD	2019	Dataset	Data from: Asymmetrical habitat coupling of an aquatic predator – the importance of individual specialisation	Dryad	0	1	0	2019-01-22T19:26:06Z	2025-07-22T06:36:33Z
☐	10.15468/ddb99b	Global Biodiversity I...	2019	Dataset	Bacterial in the Amundsen Sea Polynya (Southern Ocean): community composition in environmental samples...	SCAR - Microbial A...	0	1	0	2019-01-23T13:39:14Z	2025-09-24T12:03:11Z
☐	10.7910/dvn/fd8o8w	Harvard Dataverse	2019	Dataset	Replication Data for: Protection Through Presence: UN Peacekeeping and the Costs of Targeting Civilians	Harvard Dataverse	0	1	0	2019-01-28T19:48:36Z	2025-06-03T12:34:53Z
☐	10.5061/dryad.fn38f87	DRYAD	2019	Dataset	Data from: Postglacial colonization routes coincide with a life history breakpoint along a latitudinal gradient	Dryad	0	1	0	2019-01-29T00:16:39Z	2025-07-22T06:40:21Z
☐	10.5061/dryad.t78400r	DRYAD	2019	Dataset	Data from: A full annual perspective on sex-biased migration timing in long-distance migratory birds	Dryad	0	1	0	2019-01-31T17:51:39Z	2025-07-22T06:37:51Z
☐	10.5281/zenodo.25556...	Zenodo	2019	Dataset	Annotations of sanSuaA1 genome assembly	Zenodo	0	1	0	2019-02-02T11:08:56Z	2020-02-29T04:33:27Z
Page Size: 100 ▼ 2,001 to 2,100 of 26,873 K < Page 21 of 269 > All											



Svårigheter och svagheter

- Många aktörer har låg metadatakvalitet
- "Alla gör olika" – heterogeniteten gör det svårt att filtrera ut
- Uppdaterar inte DOI-metadata efter skapandet
- Relationer som endast syns utanför metadata
- Accessionsnummer och liknande fall – kräver text mining

Avslut

Frågor?

Projektkontakt: therese.tikkanen@chalmers.se
eller skriv i
Basecamp-tråden <http://bit.ly/4jm8Wft>



SND

Svensk nationell datatjänst | Göteborgs universitet - Chalmers tekniska högskola - Karolinska Institutet - Kungliga Tekniska högskolan - Lunds universitet - Stockholms universitet - Sveriges lantbruksuniversitet - Umeå universitet - Uppsala universitet