

Granskning av dataset hos SU

Merlijn de Smit

Datakoordinator, Forskningsdatagruppen på SUB



Stockholms
universitet

Våra repositorer

Figshare: generell repositorium, mycket annat än dataset (powerpoints, konferenspresentationer, osv.). 12 dataset i 2023, 7 hittills i 2024.

Dataverse: generell repositorium, bara dataset. Enstaka uploads (10 totalt).

Zenodo: generell repositorium, mycket annat än dataset. Enstaka uploads (37 "items" totalt varav 8 dataset).

SND: generell repositorium med områdesspecifika metadata, bara dataset. 87 SU-dataset total, 2 hittills under 2024.

Bolin Centre Database: klimatologisk repositorium för data från forskare med anknytning till Bolincentret (SU+KTH+SMHI). 432 dataset totalt, ungefär 1 per vecka.

Vem kurerar?

Innan 2024:

- 3-4 kurerare, fördelad efter veckodag.
- Avisering av nya dataset via e-post.
- Periodvis skördning av metadata och data till arkiv från **Zenodo** och **Figshare**, "halvautomatiskt" (skripts i bl.a. Bash, Oxygen, Xbase). 1 person ansvarig.

Nu:

- 2 person som arbetar med dataset för Bolincentrets databas, 1 person som arbetar med alla andra repositorer (Figshare, SND etc.).
- Bolindataset fördelas av föreståndare databas (Anders).
- Bolincentrets data kureras också av person som jobbar på Bolincentret (1-2 kurerare).
- Skördning av **Zenodo** och **Figshare** fortsätter, också igång med **Data Dryad** (där vi inte kurerar, ändå ~300 SU-dataset), inte börjat än med Bolincentrets data.



Figshare och Bolin

Dataset on the expression "Canadian model" in the Swedish Press and student papers (2021-2023)

Cite

Download all (2.7 MB)

Share

Embed

+ Collect

...

Dataset posted on 2024-09-27, 13:25 authored by [Christophe Premat](#)

This dataset is part of a study exploring representations of the "Canadian model" in Swedish media and among students enrolled in Canadian Studies courses at Stockholm University. The data covers media articles, radio transcripts, and anonymized student essays collected between 2021 and 2023. It includes qualitative and quantitative content, focusing on discourse analysis of how the "Canadian model" is portrayed in social, political, and academic contexts.

Swedish Media Corpus (2021-2023):

This section contains articles from various Swedish newspapers as well as radio transcripts that mention the "Canadian model" or similar terms such as "kanadensisk modell" and "kanadensisk förebild." The media content was sourced from the Mediearkivet database and the National Library of Sweden. The list of articles (from Mediearkivet) is present but for copyright issues, we were not allowed to attach the whole articles.

Student Papers on Canadian Studies (2021-2023):

This section includes anonymized essays written by students from Stockholm University as part of an introductory course on Canadian Studies. Each essay engages with the concept of the "Canadian model," responding to a quote from Canadian Prime Minister Justin Trudeau. The student papers are available in text format (PDF/Word) and are provided with metadata that includes the year of submission and the total word count. The collection comprises a total of 155,692 words across 87 student essays. This dataset is available for academic research purposes under the terms of a Creative

USAGE METRICS

23
views

14
downloads



CATEGORIES

• Cultural studies

KEYWORDS

Swedish Press

stereotypes

Bolin Centre Data Oden data repository Data from expedition ASCOS, Arctic Ocean, 2008

Vertical stratification of submicrometer aerosol particles measured during the high-Arctic ASCOS expedition 2008

Caroline Leck, Joseph Sedlar, Erik Swietlicki, Staffan Sjögren, Barbara Brooks, Sarah Norris

The atmospheric thermodynamic structure and aerosol particle concentrations, in several size ranges, across a multitude of surface conditions over the Arctic Ocean pack ice, was profiled using a helicopter based onboard Swedish icebreaker Oden during the Arctic Summer Cloud Ocean Study (ASCOS) expedition in summer 2008.

Aerosol sources are known to be few in summer high Arctic while the cloud cover is generally at a maximum during this season. Until now, a large deficiency has been an inability to relate Arctic aerosol and cloud properties because of the difficulties in performing in situ measurements within the harsh Arctic environment.

The ultrafine particle (diameters between 3–15 nm) concentrations were typically a few hundred per cm^3 , but reached as high as 1500 cm^{-3} and were often found within the shallow Arctic boundary layer, while occasional elevated concentrations were measured at higher altitudes. The accumulation mode particle concentrations were generally less than 1 cm^{-3} within the well-mixed boundary layer. Such low levels can be explained by the presence of a moist boundary layer combined with in general weak local sources. In addition, largest accumulation mode numbers were found at and above former cloud tops.

Atmosphere Aerosols Aerosol vertical profiling High Arctic Summer ASCOS Arctic boundary layer Arctic clouds Icebreaker Oden

Download data

Arbetsflöde: Figshare

- Avisering av ny dataset via e-post.
- Admin inloggning: kurerare kan ta ansvar för en dataset.
- Intern gränssnitt i systemet för kurering inkl. kommentar. Metadata och data kan ändras direkt (om kurerare "loggar in som" användare) men vanligtvis ber vi om ändringar via kommentarfunktionen.
- Godkännande (och DOI) direkt via "approva and publish"-knappen.

Arbetsflöde: Bolin

- Avisering av ny dataset via e-post. Arbetet fördelas och organiseras via Asana.
- Ingen admin-inloggning. Kurerare har samma gränssnitt som användare. Gränssnittet har en versioneringssystem. Kommunikation med användare via e-post.
- Vanligt att vi ändrar direkt i användargränssnittet i samråd med användare. Karta med geografiska koordinater.
- För DOI och slutpublikation behövs tillgång till specifik dator



Vad granskar vi i Figshare?

Figshare:

- Filnamn, filformat (beskrivande, tillgänglig?)
- Ger beskrivningen tillräckligt med information för att man förstår vad som finns i filer, hur de öppnas, osv.?
- Är variabler förklarade i en kodbok, readme-fil, osv.?
- Stämmer referensen?

Det finns en checklista som vi kan också dela med användaren.

Checklist for items in Stockholm University Figshare repository

File names:

must not contain any forbidden characters or white space. Only permitted character set is: A-Za-z0-9_-.

Preferably use dot (.) only once, for separation of file extension.

Filenames should if possible be descriptive, “contextual” in the sense that they could be understood also out of context, and pertinent to file content to make them more searchable and Findable (- one of the **FAIR**-principles) in a cross-disciplinary environment (such as Figshare). Here are some guides to good practice in file naming: [DataCarpentry](#), [DataOne](#), [Dryad](#), [Stanford](#)

File formats:

Preferably use non-proprietary, sustainable formats (e.g. .txt rather than .doc; .csv or .tsv rather than MS Excel .xls etc. In case non-proprietary formats do not offer the same functionality and layout options

(e.g. an Excel workbook with several sheets, embedded diagrams, images etc., if possible provide the same content in both proprietary and non-proprietary formats.)

For enhanced **Re**-usability (reproducibility) – another tenet of **FAIR**, it is important to **describe the**

Vad granskar vi i Bolincentrets databas?

I stort sätt detsamma som i Figshare men också:

- Namn av dataset (begriplig, beskrivande, annan än namn på publikation).
- Abstrakt (tre stycken om "vad", "hur", "när")
- Databeskrivningen – ska innehålla information om variabler osv., gärna snyggt formaterad i markdown.
- Själva data (överensstämmer allt med beskrivningen, uppenbara fel).

Detaljerad checklista/lathund för intern användning.

Anvisning för forskare i gränssnittet.

Check before pressing publish button

1. Internal name - is it as agreed? - and unique - and consistent with the data collection when relevant?
2. Title - is it understandable?
3. Title - is it coherent with other similar or related datasets?
4. Summary - does it have three paragraphs following the instructions?
5. Summary - should not contain references to articles, and avoid unexplained technical terms
6. Replace range hyphens
7. References - consistently formatted? Preferably as
springer-basic-author-date-no-et-al, en-GB, followed by a period and complete doi
web address in the form <https://doi.org/10.xxxx> <https://citation.crosscite.org/>

[unpublished manuscript](#) → Ferasat, Z., Panahi, R., & Mokhtarani, B. (2019). Natural polymer matrix as safe flocculant to remove turbidity from kaolin suspension: Performance and governing mechanism. [Manuscript in preparation]. Department of Chemical Engineering, University of Cambridge. doi

8. Data creator - should include all persons that shall be mentioned when the dataset is cited
9. Keywords - adequate, correctly spelt and formatted?
10. Category and Sub-category - appropriately chosen?
11. Contact information - correct in all details?
12. Data description - clear and understandable?
13. GCMD keywords and location - correctly chosen?



Skillnader

Figshare (och SND, Zenodo, osv.):

- Många olika ämnesområden.
- Beskrivningen ska "funka" för dataset, men i övrigt inga uttalade krav.
- Vi kollar om filer finns, om de går att öppna, men vi inte hårdgranskar själva data.

Bolincentrets databas:

- Ett ämnesområde (klimatologi).

Därför också:

- Krav på mer specifik information som ska finnas med i abstraktet (t.ex. relevant tidsperiod), koordinater som används för karta, osv.
- Krav på mer specifik databeskrivning med t.ex. information om variabler.
- Kolla om själva data överensstämmer med databeskrivning, om det finns uppenbara fel, osv.

Utmaningar

Personuppgifter: kan inte publiceras direkt i någon av våra repositorer (och vi gör detta tydligt för användare). Metadatapost istället.

- **SND:** interimslösning för publicering av (känsliga) PU via Sunet Drive nu på plats, 5 dataset.
- **Bolin:** arbetar på avtalstext med tydlig information om sekretessbelagda uppgifter.

Stora data: har inte varit en problem för Figshare, SND osv (dock: nyligen förfrågan om tillgängliggörandet av 200TB dataset!). Vi har fått in ansökningar om lagring av större datamängder på Bolincentrets databas som vi tillsvidare bedömer från fall till fall (tekniska möjligheter finns, frågan är kostnader).