

# Flaggskeppsprojekt Bolincentrets Databas

Merlijn de Smit

Forskningsdatagruppen på Stockholms universitetsbibliotek

[merlijn.de.smit@su.se](mailto:merlijn.de.smit@su.se)



Stockholms  
universitet

# Vad är Bolincentret?

Ett [tvärvetenskapligt konsortium](#) av SU, KTH och SMHI för forskning och forskarutbildning inom klimatvetenskap.

- Ung. 400 anslutna forskare
- Forskningsteman: klimatets fysik och kemi; vatten, biokemi och klimat; klimathistorik samt klimat, ekosystem och biodiversitet
- Utbildning på master- och doktorandnivå
- Konferenser och öppna föreläsningar
- Bolincentrets databas för forskningsdata



The screenshot shows the website for the Bolin Centre for Climate Research. The header includes the name of the center and navigation icons for search, menu, and a globe. The main content area features a 'CALENDAR' section with a calendar icon. Two events are listed: one for May 14th, a seminar by Wojtek Grabowski from NCAR, USA, titled 'Broadening of cloud droplet spectra through eddy hopping: Why did we all have it wrong?'; and another for May 15-17th, the 'Swedish Climate Symposium 2024'. To the left of the calendar, there are two article teasers: one from the Department of Geological Sciences about 'Atmospheric Melt 2023' led by Julia, and another about 'Arctic studies' with a play button icon.

# Vad är Bolincentrets databas?

[Ett repositorium](#) för klimat- och jordvetenskapliga data med i nuläget 403 dataset.

- Öppen för forskare med anknytning till Bolincentret, i de allra flesta fall SU-forskare
- Kartor och andra möjligheter till visualisering av data
- Ämnesspecifika metadata, kurering, DOI
- Tematiska samlingar (t.ex. isbrytare Odens data)
- I framtiden: "efterhandskurering" av dataset, hantera stora dataset (>1 TB), kodrepositorium Git.

Bolin Centre > Data > Oden data repository > Data from expedition Arctic Ocean, 2018

## Aerosol chemical composition during the Arctic Ocean 2018 expedition

Lubna Dada, Julia Schmale, Kaspar Daellenbach, Andrea Baccharini

This data set provides the bulk aerosol chemical composition measured with a high-resolution Aerosol Mass Spectrometer (HR-AMS) during the Arctic Ocean 2018 expedition. Chemical species are organics, sulfate, nitrate, sodium and ammonium. The data refers to submicron particles which can evaporate at 600 °C. The instrument was placed on the 4th deck of icebreaker Oden and sampled total as well as interstitial aerosol.

Measurements were performed onboard Swedish icebreaker Oden during August and September 2018 along the track of the Arctic Ocean 2018 expedition in the central Arctic Ocean. These data can be used to study aerosol-cloud interactions in the high Arctic.

Atmosphere Aerosols Aerosol chemical properties Aerosol chemical composition High Arctic Arctic Arctic Ocean  
Arctic Ocean 2018 AO18 AO2018 Icebreaker Oden

Download data

# Flaggskeppsprojekt ”Certifiering av Bolincentrets databas”

I nuläget: databasen drivs med mycket begränsade resurser (två medarbetare på deltid, infrastruktur på lokal server, data och metadata på Sunet).

**Mål:** att säkerställa databasens långsiktiga framtid genom att flytta infrastrukturen till mer central lösning och integrera det dagliga arbetet med databasen (kurering, utveckling, osv.) i SUB:s forskningsdatagruppens arbete.

**Ramverk för detta:** att förbereda en ansökan om **certifiering hos CoreTrustSeal** av databasen.

Detta innebär att:

- Personaltillgång måste säkras. Kompetensdelning, dokumentation av tekniska infrastrukturen
- Infrastrukturen måste vara hållbar (flyttas mer centralt, SU-IT eller Sunet)
- Andra saker: avtalstext inför publicering, utveckling metadata, rutiner för arkivering och avveckling

# Projektgrupp och styrgrupp

## Projektgrupp:

- Anders Moberg (koordinator databas)
- Rezwan Mohammad (teknisk drift och utveckling)
- Joakim Philipson (metadataspecialist)
- Maria Almbro (kurering)
- Julia Baumann (kontakt med SU-IT, flytt)
- Merlijn de Smit (projektledare)

Under vintern och våren har vi också haft hjälp av Philip Süess med kurering av inkommande dataset.

## Styrgrupp:

- Wilhelm Widmark (överbibliotekarie)
- Mia Wahlberg (sektionschef SUB)
- Gustaf Hugelius (föreståndare Bolincentret)
- Iris Alfredsson (SND)



# CTS-ansökan, hur är läget?

- Mycket är redan på plats, eller är på plats men ska dokumenteras. T.ex.: R2 "Rights management" och R4 "Legal and ethical": vi behöver en avtalstext mellan forskaren och databasen med intyg att data inte innehåller känsliga uppgifter.

Det som är inte på plats än och är prioriterad är till exempel:

- R3: Continuity of service. Säkra långsiktig personaltillgång, långsiktig drift men också planer för överlämning eller avveckling om driften skulle bli omöjlig.
- R14: Storage and security. Dokumentera hur data och metadata skyddas från obehörig förändring eller förlust. Förutsätter flytt av infrastrukturen till mer central lösning.
- R15: Technical infrastructure. Databasens infrastruktur ska dokumenteras, tillgänglighet ska garanteras och det ska dokumenteras hur infrastrukturen kan utvecklas efter behov.
- R16: Security. Beslut och dokumentation om samarbete med Sunet och SafeSpring behövs.



# Databasens infrastruktur

## Nuvarande infrastruktur:

- Data och metadata på Sunet, databasens infrastruktur på lokal server på SU.
- PHP
- Kureringsgränssnitt med begränsad funktionalitet. Kurering via e-postkommunikation, publicering av metadata på lokal maskin.

## Nya infrastrukturen:

- Data och metadata på Sunet, databasens infrastruktur implementeras via en Kubernetes-lösning på mer central server (SU-IT, Sunet, extern leverantör). Infrastrukturen underhålls/utvecklas via GitOps.
- NodeJS
- Inloggning för olika roller (kurerare, forskare, granskare). Kommunikation via kureringsgränssnitt.

# Flaskhalsen

Implementering av nya infrastrukturen är inte i sig en del av flaggskeppsprojektet, men mycket förutsätts av att den införs, t.ex.:

- Utveckling av metadata och skördning av metadata till [researchdata.se](https://researchdata.se).
- Mycket av det som ska dokumenteras handlar om nya systemet, inte det gamla.
- Gamla systemet förutsätter 1-2 kurerare som har tillgång till specifik dator. Kurering och arbetsflöde mycket enklare med nya systemet.





# Aktiviteter: kurering

## Utmaningar:

- Gamla systemet lämpad för 1-2 kurerare med tillgång till specifik maskin (som krävs för publicering). Kommunikation via e-post har känts rörigt. Allt detta kommer bli lättare med nya systemet.
- Kurering för (ämnesspecifika) Bolincentrets databas har varit mer hands-on och tidskrävande än kurering in de (generella) repositorer som FD-gruppen har hittills tagit hand om.

## Nuläget:

- Det kommer in (ungefär) 50 dataset per år, dvs. en per vecka.
- En kurerare från Forskningsdatagruppen har fått tillgång till datorn och databasens system för arbetsflöde. Flera från FD-gruppen kommer bli inblandad framöver.
- Det tog lite tid att komma igång, men folk från FD-gruppen är nu med i kureringen!



# Aktiviteter: avtalstext

Behov av en avtalstext mellan forskaren och Bolincentrets databas för att till exempel:

- Punkt R4 "Legal and ethical" i CTS-ansökan
- Intyg att data kan publiceras öppet och inte innehåller personuppgifter eller sekretessbelagda data
- Ange Stockholms universitet som avtalspartner
- Anvisningar för hantering eventuella konflikter mellan forskaren och repositoret.

Nuläget:

- En första utkast under arbete (med hjälp av befintliga avtalstexter SND och Data Dryad), diskussion med jurister SU och jurister SND påbörjas.

# Aktiviteter: migrering till central IT

Nya infrastrukturen baserad på kubernetes-miljö och GitOps. SU-IT planerar att bygga upp kubernetes-miljö men den finns inte på plats än. Avtal med extern leverantör som vi kunde ha använt sägs upp. Alternativ:

- Flytta gamla systemet till virtuell maskin SU-IT eller Sunet tills kubernetes-miljö på SU-IT finns på plats.
- Flytta till kubernetes-miljö extern leverantör (Amazon, Microsoft Azure), antingen direkt eller via SU-IT, som mellanlösning.
- Använda en kubernetes-lösning via Sunet.

Problem:

- Mycket hänger på att den nya infrastrukturen driftsätts.
- Kostnader! Infrastrukturen kräver inte alls mycket resurser, lösningen ska vara skalbar.
- Gamla metadata ska överföras till det nya systemet. När? Ska de två system köras parallellt ett tag?

# Aktiviteter: metadata och dokumentation

Vi **kan** göra en CTS-ansökan med det gamla systemet, men:

- CTS-ansökan är ett medel för att uppnå ett mål (långsiktig hållbarhet av databasen). Om vi CTS-certifierar databasen med det gamla systemet måste mycket relevant arbete (flytt, dokumentation, arbetsflöde) göras om senare.

Vi vill också att Bolincentrets databasens metadata kan skördas till [researchdata.se](https://researchdata.se), men:

- Det skulle också vara metadata från det nya systemet.

Nuläget:

Joakim bevakar utvecklingen i researchdata-arbetsgruppen om metadatastandarder och skördning. Men mycket arbete förutsätts av att nya infrastrukturen är på plats. Detsamma gäller dokumentation.

# Problem och lösningar

## Problem:

- Flaskhalsen med nya infrastrukturen.  
Kommunikation med SU-IT har inte alltid gått så smidigt, och mycket hänger på att flytt av infrastrukturen blir av. Det känns ibland frustrerande att vissa aktiviteter (t.ex. utveckling metadatafält och dokumentation) måste vänta.
- Andra personalrelaterade situationer har också lett till viss fördröjning.

## Lösningar:

- Vi koncentrerar oss på det vi kan göra nu.  
Kompetensdelning och samarbete när det gäller t.ex. kurering. Detta är nu på gång.
- IT-relaterad arbete går framåt, om än långsamt (två steg framåt, ett steg tillbaka). Kommunikation med SU-IT fungerar bra nu.
- Förutsättning av hela arbetet med långsiktig hållbarhet Bolincentrets databas är att databas blir en del av FD-gruppens dagliga arbete. Detta är fallet.



# Till slut

- Vara sköldpadda, inte hare. Gediget arbete och stadig framgång bättre än snabba resultat som håller kanske inte i längden.
- Blicken på målet: att säkra databasens långsiktiga framtid. Detta kräver en del "mjuka" resultat (kompetensdelning, personaltillgång på lång sikt, kommunikationskanaler mellan olika intressenter), inte bara "hårda" resultat (tekniska lösningar och dokument).

