

Publicera och dela forskningsdata - Utmaningar och möjligheter

Medicinska bibliotekskonferensen MEDBIB2022
Solna 2022-06-16

Helena Eckerbom
Glenn Haya
Karin Widin

Data Access Unit (DAU)
Karolinska Institutet University Library (KIB)
kib@ki.se

MÅLET

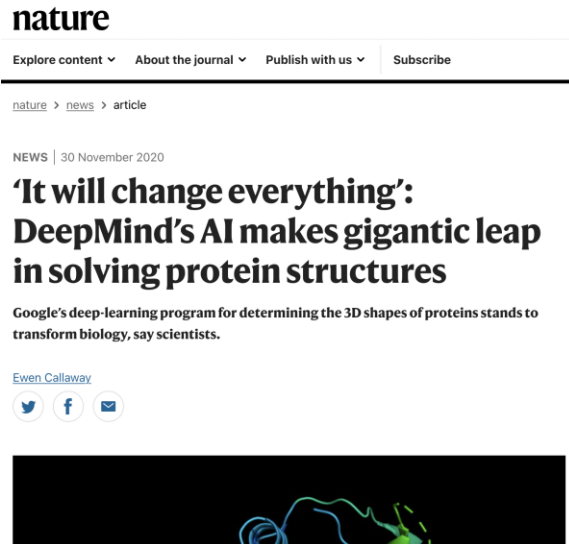
Open Data & Open Access

in the broader context of

Open Science



Open data kan leda till stora framsteg



The image shows a screenshot of a news article from the journal Nature. The article title is "'It will change everything': DeepMind's AI makes gigantic leap in solving protein structures". The byline is "Ewen Callaway". Below the text is a small image showing a protein structure rendered in blue and green. The article text mentions "Google's deep-learning program for determining the 3D shapes of proteins stands to transform biology, say scientists."

nature

Explore content ▾ About the journal ▾ Publish with us ▾ Subscribe




nature > news > article


NEWS | 30 November 2020

**'It will change everything':
DeepMind's AI makes gigantic leap
in solving protein structures**

Google's deep-learning program for determining the 3D shapes of proteins stands to transform biology, say scientists.

Ewen Callaway



"This will change medicine. It will change research. It will change bioengineering. It will change everything"

Andrei Lupas, an evolutionary biologist at the Max Planck Institute

"It's a breakthrough of the first order, certainly one of the most significant scientific results of my lifetime."

Mohammed AlQuraishi, a computational biologist at Columbia University

Vår del - hjälper forskare publicera data



[Find Data](#) [Manage Data](#) [Describe & Share Data](#) [News & Events](#) [About Us](#) [Contact](#)

[← Back to search results](#)

Agilent gene expression microarray raw data for STO-3

SND-ID: 2021-340

[Description](#) [Data and documentation](#)

[Request data](#)

Dataset

Agilent gene expression microarray raw data for STO-3

[Associated documentation](#)
List_of_probes_STO3_Agilent_microarray_gMeanSignals_652.txt

Description

The dataset consists of a table in one (1) txt file containing: Gene expression data (Agilent custom 32K microarray) generated in 2014 for 652 primary tumors from the STO-3 trial. The variables are the 31195 probes included in the Agilent microarray used to generate the data.
File name: STO3_Agilent_microarray_gMeanSignals_652.txt

The 31195 probes included in the Agilent microarray used to generate the data are listed and mapped to gene names in an annotation file.

Annotation file name: List_

...

[Show more..](#)

Version 1

2022-03-02

<https://doi.org/10.5878/3xxa-3c28>

Accessibility level
Access to data through SND
Access to data is restricted

Use of data
Things to consider when using data
shared through SND

Contacts for questions about the c
Annelie Johansson
Linda Lindström

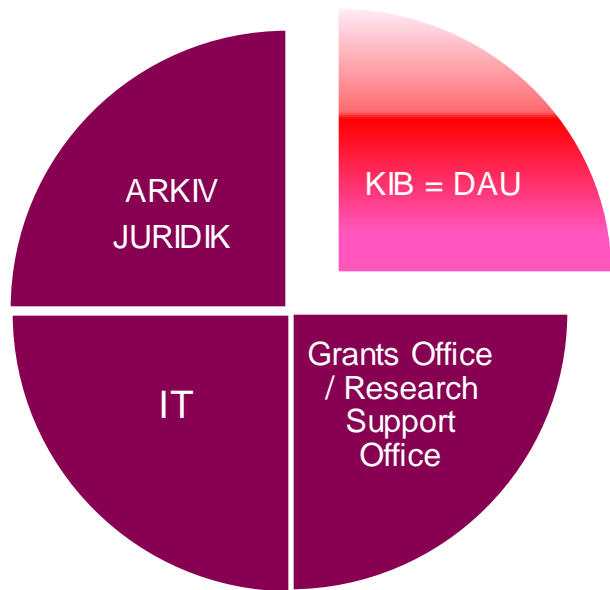
Forskningsdata och din tjänst

<https://www.menti.com>

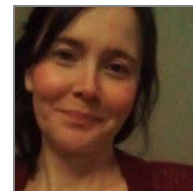
Kod: 6787 0506

VILKA ÄR VI ?

Research Data Office (RDO)



Data Access Unit (DAU)



Data Access Unit

The Data Access Unit at the University Library helps you making research data available and open access where applicable. We're part of [Research Data Office \(RDO\)](#) at KI.

Telephone: [08-52484000](tel:08-52484000)

E-mail: rdo@ki.se



Publication support

Contact us with questions regarding open access, bibliometrics, literature searching, publishing in KI Open Archive and strategic publishing.

Telephone: [08-52484000](tel:08-52484000)

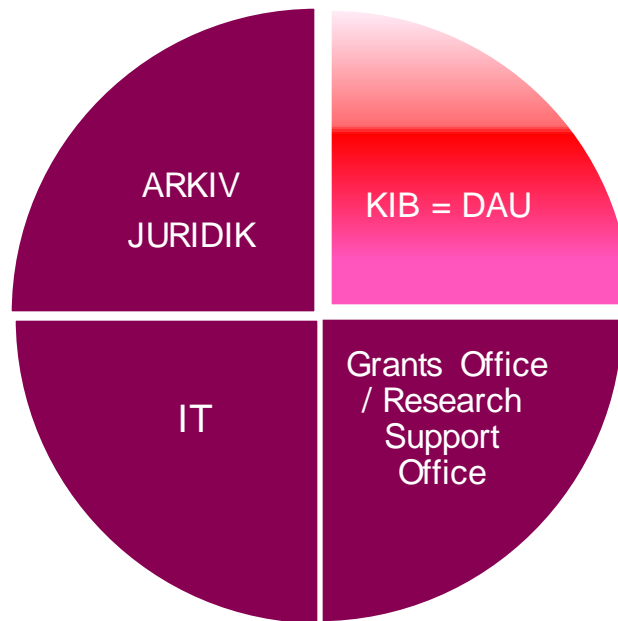
E-mail: kib@ki.se

[Book publication support](#)

RDO:s olika delar och ansvarsområden

- Arkivering forskningshandlingar
- Diarieföring, gallring
- Lagar & Regler
- GDPR – Dataskyddsombud
- Utlämningsförfrågningar av forskningsdata
- Stöd i sekretessbedömningar

- Lagring
- Säkerhet
- Informationsklassning
- S3 lagring – i samarbete med KIB DAU



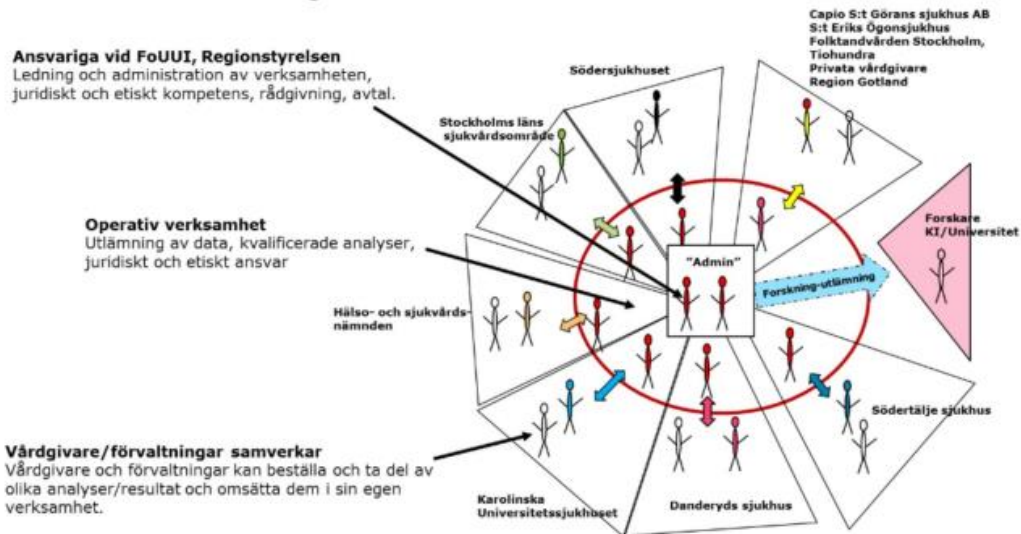
- Tillgängliggöra =
Publicera / dela forskningsdata
- Rådgivning och support
vetenskaplig publicering

- Ansökningar + återrapportering
finansiärer
- Dokumentation – Electronic
Notebook (ELN)
- Datahanteringsplaner
- Råd humanetikansökan
- Kliniska provningar

I KI:s nära omvärld...

CLINICUM

Organisation av Centrum för hälsodata



SND-konsortiet



SND:s verksamhetsmodell bygger på en distribuerad verksamhet. SND-kontoret, med stöd av de nio konsortielärosätena, fungerar som nav för de stödfunktioner för forskningsdata (DAU:er) som etablerar sig hos våra nätverksmedlemmar.

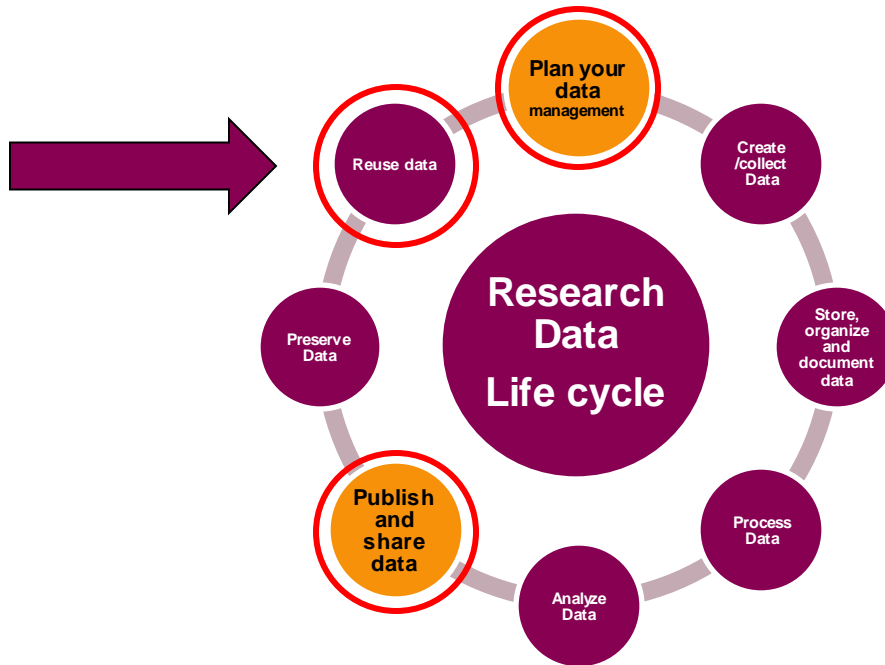
För närvarande ingår följande i SND-nätverket:

- [Blekinge tekniska högskola](#)
- [Chalmers tekniska högskola](#)
- Enskilda Högskolan Stockholm
- Försvarshögskolan
- Gymnastik- och idrottshögskolan
- [Göteborgs universitet](#)
- [Handelshögskolan i Stockholm](#)
- [Högskolan Dalarna](#)
- [Högskolan i Borås](#)
- Högskolan i Gävle
- [Högskolan i Halmstad](#)
- Högskolan Kristianstad
- Högskolan i Skövde
- [Högskolan Väst](#)
- [Jönköping University](#)
- [Karlstads universitet](#)
- [Karolinska institutet](#)
- [Kungliga tekniska högskolan](#)
- [Linköpings universitet](#)
- [Linnéuniversitetet](#)
- [Luleå tekniska universitet](#)
- [Lunds universitet](#)
- [Malmö universitet](#)
- [Mittuniversitetet](#)
- [Mälardalens högskola](#)
- Polarforskningssektariatet
- RISE Research Institutes of Sweden
- Sophiahemmet Högskola
- Statens veterinärmedicinska anstalt
- [Stockholms universitet](#)
- [Sveriges lantbruksuniversitet](#)
- [Södertörns högskola](#)
- [Umeå universitet](#)
- [Uppsala universitet](#)
- [Örebro universitet](#)



Forskningsdatalivscykeln

- med fokus på tillgängliggörande



PUBLICERINGSLANDSKAPET

Finansiärernas och förlagens krav



EUROPEAN UNION

HORIZON EUROPE

#HorizonEU

14 JUN 2021

"Horizon Europe will set a new standard for dissemination of knowledge and new skills across European societies. With clear and immediate open access requirements for beneficiaries, the Open Europe Research publishing platform and a strengthened European Open Science Cloud, we are well underway in making truly open science a reality!"

Margrethe Vestager, Commissioner for Innovation, Research, Culture, Education and Youth

OPEN SCIENCE

EARLY KNOWLEDGE AND DATA SHARING, AND OPEN COLLABORATION

The challenge is for Europe to embrace open science as the modus operandi for all researchers. Open science consists in the sharing of knowledge, data and tools as early as possible in the Research and Innovation (R&I) process, in open collaboration with all relevant knowledge actors, including academia, industry, public authorities, and users, citizens and society at large. Open science has the potential to increase the quality, efficiency and impact of R&I, lead to greater responsiveness to societal challenges, and increase trust of society in the science system.

What are open science practices?

- Open access to research outputs such as publications, data, software, models, algorithms, and workflows;
- Early and open sharing of research, for example through preprintation, registered reports, pre-prints, and crowd-sourcing of solutions to a specific problem;
- Use of open research infrastructures for knowledge and data sharing;
- Participation in open peer review;
- Measures to ensure reproducibility of results; and
- Open collaboration within science and with other knowledge actors, including involving citizens, civil society and end-users, such as in citizen science.



Vetenskapsrådet

Samordningsuppdrag om öppen tillgång till forskningsdata 2022

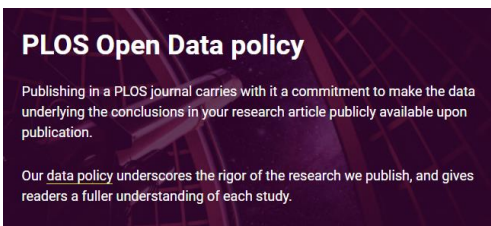
Vetenskapsrådet 2022



Open Research Europe

Publish any research outputs you wish to share across all fields of science and technology.

European Commission | Powered by F1000 Research



PLOS Open Data policy

Publishing in a PLOS journal carries with it a commitment to make the data underlying the conclusions in your research article publicly available upon publication.

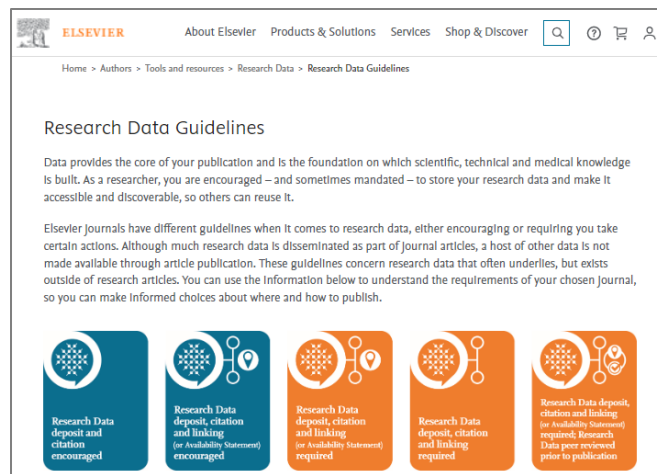
Our data policy underscores the rigor of the research we publish, and gives readers a fuller understanding of each study.

Data availability statement (DAS)

Authors are required to provide a data availability statement, detailing where data associated with a paper can be found and how it can be accessed. The DAS should be submitted within the article manuscript, before the 'References' section. If data cannot be made open, authors should state why in the data availability statement.

The table below contains some template statements which comply with our open and FAIR policy. This is not an exhaustive list however, and an individual data set might warrant a different type of statement.

Availability of data	Template for data availability statement
Data openly available in a public repository (that supports FAIR principles) that issues datasets with DOIs under open licence (e.g. CC0)	The data that support the findings of this study are openly available in [repository name e.g. "figshare"] at http://doi.org/[doi] , reference number [reference number].



ELSEVIER

About Elsevier | Products & Solutions | Services | Shop & Discover

Home > Authors > Tools and resources > Research Data > Research Data Guidelines

Research Data Guidelines

Data provides the core of your publication and is the foundation on which scientific, technical and medical knowledge is built. As a researcher, you are encouraged – and sometimes mandated – to store your research data and make it accessible and discoverable, so others can reuse it.

Elsevier journals have different guidelines when it comes to research data, either encouraging or requiring you take certain actions. Although much research data is disseminated as part of journal articles, a host of other data is not made available through article publication. These guidelines concern research data that often underlies, but exists outside of research articles. You can use the information below to understand the requirements of your chosen journal, so you can make informed choices about where and how to publish.

Research Data deposit and citation encouraged

Research Data deposit, citation and linking (or Availability Statement) encouraged

Research Data deposit, citation and linking (or Availability Statement) required

Research Data deposit, citation and linking required

Research Data deposit, citation and linking (or Availability Statement) required; Research Data peer reviewed prior to publication

Publiceringslandskapet forts.

- FAIR data principer har tagits fram som en målbild vi vill uppnå
- Känsliga data och GDPR skapar många komplexa frågor

VÅRT ERBJUDANDE

Data Access Unit

Sedan hösten 2020 finns en nyinrättad grupp, Data Access Unit (DAU) på Karolinska Institutets universitetsbibliotek (KIB). Vi är ett team med olika kompetenser som har till uppgift att stödja dig som forskare och medarbetare på KI i processen med att publicera och dela dina forskningsdata.

DAU har ett nära samarbete med universitetsbibliotekets [Publiceringsstöd](#). Tillsammans kan vi ge dig ett stöd genom hela den vetenskapliga publiceringsprocessen.

DAU-gruppen ansvarar specifikt inom [Research Data Office \(RDO\)](#) för att publicera och dela KI:s dataset, både med fullt öppen tillgång och begränsad åtkomst. DAU hjälper till vid publicering av databeskrivningar i Svensk Nationell Datatjänsts (SND) katalog. I detta ingår en granskning de beskrivningar som skapas via SND:s dokumentationssystem [DORIS](#).

Publicering och delning av forskningsdata underlättar återanvändning och reproducerbarhet.

SND:s forskningsdatakatalog

DORIS

DORIS
SND:s datagransnings- och informationsystem

Databeskrivningar | Förfrågningar om data | Administration | Mina sidor | Logga ut

Inledande uppgifter

Titel *
Ange en titel för din databeskrivning, både på svenska och engelska. Tänk på att det är viktigt med en tydlig titel som ger information om vad det är för data som beträffas.

Det är möjligt att använda en titel som redan finns i ett annat sammanhang, exempelvis ett projektnamn eller titel på en artikel. Om du använder en redan befintlig titel som inte kan identifieras för svenska kan den engelska titeln anges. (Både titeln, tillägget och specifikationerna för databeskrivningen.)
Använd S. Data, Inledning & Årsmat

Titel (på svenska)

Titel (på engelska)

Tillgänglighetsnivå för data *

Tillgänglig till data via SND eller annan ekter *
Ange var data kommer från tillgängliga. Du väljer om databeskrivningen ska göras tillgänglig via SND:s forskningsdatakatalog, genom meddelande eller via en förfrågan, eller via:
 Rekommendation till data via SND
 Rekommendation till data via extern ekter

Nivå av tillgänglighet *
Data som delas via SND kan ha en av två nivåer: [öppen tillgänglighet](#). Detta bör enligt [SND:s principer för öppen tillgänglighet](#) vara kopierbar till andra om det är tillgängligt. [Läs mer](#)
 Data är fritts tillgänglig
 Tillgänglig till data är begränsad

Data innehåller personuppgifter *
Här anges om data innehåller personuppgifter. Med detta menas:

Sök i SND:s katalog SÖK → ⓘ

Aktiva filter: Karolinska Institutet X

32 träffar

< Föregående 1 2 3 4 Nästa >

CSAW-CC (mammografi) – ett dataset för AI-forskning för att förbättra screening, diagnostik och prognostik för bröstcancer Karolinska Institutet

Detta dataset innehåller röntgenbilder, mammografi, från bröstcancerscreening på Karolinska Universitetssjukhuset för perioden november 2008 till december 2015. Datasetet har sammanställts med syftet att utföra AI-forskning...

- Fredrik Strand, Karolinska Institutet, Institutionen för Onkologi-Patologi

Publicerad: 2022-04-22 ↻

S3

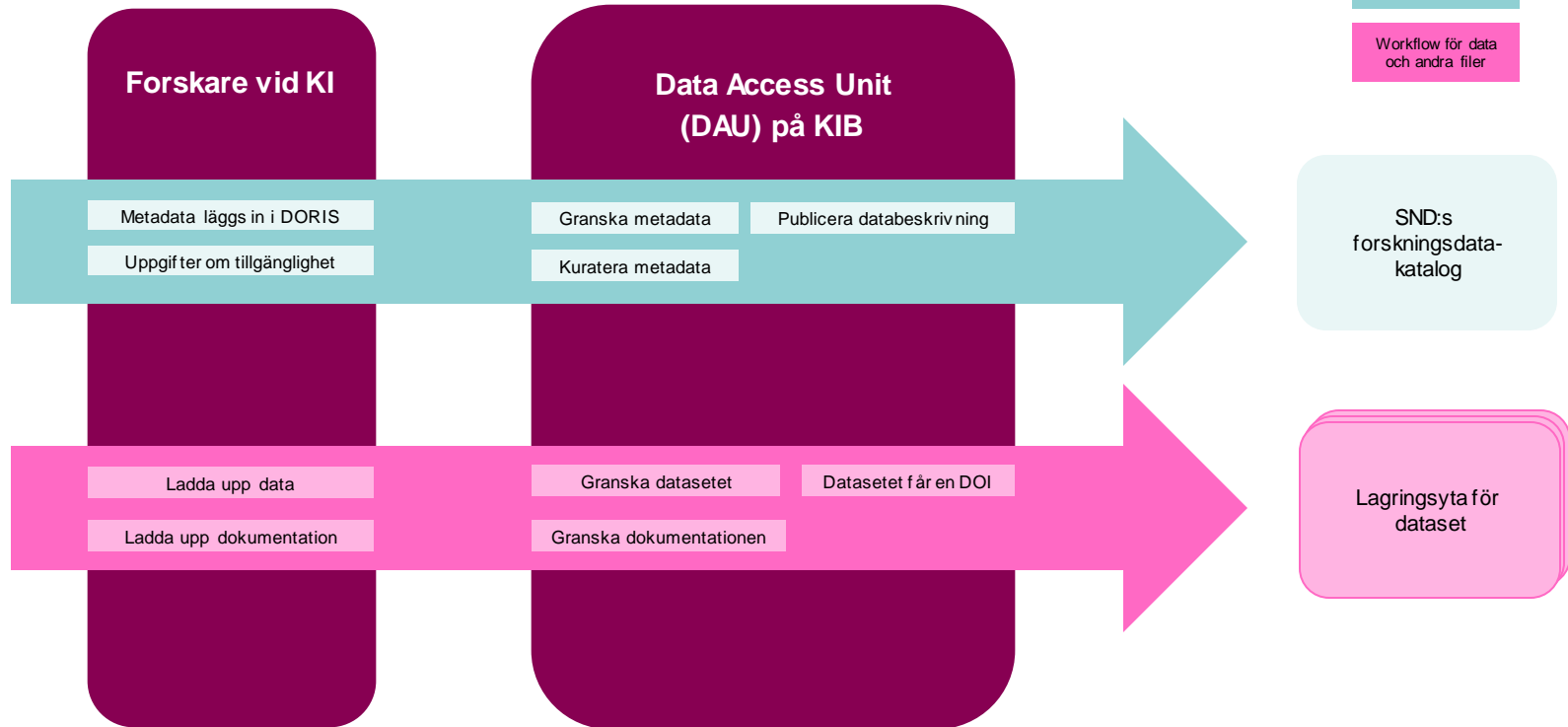
data

Ladda ner | Redigera | Egenskaper | Ny | + | -

/2021-204/2021-204-1-1/data/

Namn	Storlek	Ändrad
anon_annotations_nonhidden		
00002_20990909_L_CC_1.dcm	16 641 KB	2022-04-13 13:22:46
00002_20990909_L_MLO_1.dcm	16 641 KB	2022-04-13 13:22:46
00002_20990909_R_CC_1.dcm	16 641 KB	2022-04-13 13:22:46
00002_20990909_R_MLO_1.dcm	16 641 KB	2022-04-13 13:22:46
00004_20990909_L_CC_1.dcm	26 625 KB	2022-04-13 13:22:46
00004_20990909_L_CC_2.dcm	16 641 KB	2022-04-13 13:22:46
00004_20990909_L_MLO_1.dcm	16 641 KB	2022-04-13 13:22:46
00004_20990909_L_MLO_2.dcm	26 625 KB	2022-04-13 13:22:46
00004_20990909_R_CC_1.dcm	26 625 KB	2022-04-13 13:22:46
00004_20990909_R_CC_2.dcm	16 641 KB	2022-04-13 13:22:46
00004_20990909_R_MLO_1.dcm	16 641 KB	2022-04-13 13:22:46
00004_20990909_R_MLO_2.dcm	26 625 KB	2022-04-13 13:22:46
00005_20990909_L_CC_1.dcm	26 625 KB	2022-04-13 13:22:46
00005_20990909_L_CC_2.dcm	26 625 KB	2022-04-13 13:22:46
00005_20990909_L_CC_3.dcm	26 625 KB	2022-04-13 13:22:46
00005_20990909_L_CC_4.dcm	26 625 KB	2022-04-13 13:22:46
00005_20990909_L_MLO_1.dcm	26 625 KB	2022-04-13 13:22:46
00005_20990909_L_MLO_2.dcm	26 625 KB	2022-04-13 13:22:46
00005_20990909_L_MLO_3.dcm	26 625 KB	2022-04-13 13:22:46

DORIS & S3



Utbildning + webbsidor

Kurskatalog forskarutbildning - HT22

Startsida > Ansökan kan ske mellan 2022-04-19 och 2022-05-16

Meny	Application closed
Index	Skriv ut
Samtliga kurser	Titel Open science and reproducible research
Lista över kurser per program	Kursnummer 2963
Lista över kurser per institution	Program 0-Infle del av forskarutbildningsprogram
Behörighet och urval	Språk Engelska
Allmän datakvalitetsförordning	Antal högskolepoäng 3.0
Katalog i PDF (VT22)	Datum 2021-03-08 -- 2021-03-19
Katalog i PDF (HT22)	
Kursansökan	Kursansvarig institution Institutionen för klinisk neurovetenskap

[I www.kurser.ki.se](#)

Workshop: Publish and share your data

Do you need to publish your data? In this workshop, you will get an overview of what to do. We will cover the process of how to publish your data if it will be available as open access, as well as what to do if you have sensitive data which cannot be made openly available. We will also discuss your options regarding where to store and share your data.

19 May

13:00-15:00
CET

Online
Zoom

🏠 / Kalender / Webinar: Open access to scientific publications and research data in Horizon Europe and H2020

Webinar: Open access to scientific publications and research data in Horizon Europe and H2020

18 November

13:00-14:00
CET

Online
Zoom

[Se hela kalendern →](#)

During this webinar you will learn about open science and open access publishing of scientific publications and research data, with a specific focus on the Horizon Europe and H2020 programmes. The new EU programme Horizon Europe contains a number of requirements regarding open science and open access that KI researchers need to follow and know about when applying for research funding.

Publicera och dela forskningsdata

Bibliotekets Data Access Unit (DAU) hjälper dig med hur man ska tänka och agera kring publicering av forskningsdata.

Support

Hantera forskningsdata i 5 steg

Väg

Publicera och dela forskningsdata - en översikt

Genom publicering av data görs det möjligt att efterhand verifiera publicerade resultat eller utreda vetenskaplig oredlighet. Data som tillgängliggörs på ett rekommenderat och standardiserat sätt blir väl synliggjorda, sökbara och återbara.

DORIS - publicera dina forskningsdata

Svensk Nationell Datajäms (SND) har ett system för att beskriva och dela forskningsdata kallat DORIS. Med hjälp av DORIS kan forskare publicera uppgifter om dataset oavsett om de kan göras öppet tillgängliga eller om de innehåller känsliga data, till exempel personuppgifter.

FAIR-principerna

Enligt KIs open access-policy ska publikationer, när så är möjligt, uppfylla FAIR-principerna (Findable, Accessible, Interoperable, Reusable).

Finanslärares & förlägnars krav

I Europe och USA ställs allt oftare krav på att man som forskare samtidigt ska publicera både vetenskaplig tidsskriftsartiklar och öppen data som ligger till grund för forskningsresultatet.

Datarepositorier

För att tillgängliggöra och publicera dina data öppet rekommenderas att registrera dem i ett datarepositorium. På så sätt tilldelas även ditt dataset ett unikt ID, en permanent identifierare även kallad PID, som kan hänvisas till ett eller flera objekt.

[Mer information om forskningsdata finns på KIs medarbetarsportal →](#)

KI:s workshoppar om forskningsdata

Just nu har vi i gång evenemang inom detta område.

[Se alla evenemang →](#)

Beställ en workshop om forskningsdata

Är du intresserad av publicering av forskningsdata? Data Access Unit kan hjälpa till med att ordna workshoppar för forskargrupper/institutioner och enheter. [Klicka in på intresseanmälan](#) på återkopierad länk.

[Kontakta oss](#)

UTMANINGAR

Innehåller data personuppgifter?

- Anonymiserade eller pseudonymiserade data
- Vad gäller för aktuell datatyp?

Data innehåller personuppgifter *

Här anges om data innehåller personuppgifter. Med detta menas:

- **Direkta identifierare.** Gäller t.ex. namn, personnummer, fotografier av personer eller adresser.
- **Indirekta identifierare.** Kan vara uppgifter som i kombination med varandra möjliggör identifiering, exempelvis uppgifter om exakt inkomst, yrke och postnummer.
- **Pseudonymiserade data.** Inga identifierare finns i datamaterialet, men det finns relaterad information som kan möjliggöra identifiering av individer (t.ex. via en **kodnyckel** som finns bevarad, oavsett var kodnyckeln finns).

EU:s dataskyddsförordning (GDPR) reglerar bland annat om och hur personuppgifter får behandlas för forskningsändamål. Enligt GDPR är en personuppgift all slags information som direkt eller indirekt kan hänföras till en fysisk person som är i livet.

Data som är *avidentifierade* omfattas inte av EU:s dataskyddsförordning. För att avidentifiera personuppgifter måste alla möjligheter till identifiering vara borttagna, så att uppgifter i data inte längre går att koppla till en fysisk person som är i livet. Detta innebär att kod- och krypteringsnyckel inte finns kvar och att bakvägsidentifiering inte är möjlig.

[På SND:s hemsida kan du läsa mer om forskningsdata med personuppgifter.](#)

[Visa mindre](#)

Ja Nej

Var ska forskningsdata publiceras?



DORIS →

In DORIS, you can create and manage data descriptions that will be made accessible in the SND research data catalogue. You can also follow and manage requests for data.

GitHub



Var ska forskningsdata publiceras?

- Är det känsliga data?
- Finns det krav från förlag eller finansiär om repositorium?
- Var publicerar andra forskare forskningsdata?
- Persistenta identifierare (PID)
- Certifierat repositorium
- Avgifter
- Avtal
- FAIR

FAIR – höga krav på metadata

- Following FAIR data principles means your data can be found and reused by other people as well as machines



Gränser för vårt arbete och vem betalar?



- Granskning
- Kuratering
- Lagring

Begränsad tillgång till känsliga data



- ⑩ Autentisering
- ⑩ Säkerhet
- ⑩ Provning
- ⑩ Avtal

Stora dataset



- ⑩ Svårt att flytta
- ⑩ Svåröverskådlig
- ⑩ Nya versioner vid förändring

FRAMTIDEN

Vi tror att...

- Publiceringen av forskningsdata kommer att normaliseras
 - Kopplingen mellan artikel och tillhörande forskningsdata kommer att stärkas
 - Mängden av forskningsdata som publiceras kommer att öka
 - Publicering av forskningsdata kommer behöva ingå i planeringen från början
 - Verktyg för att publicera och hitta forskningsdata kommer att förbättras
- Vi kommer att bli bättre på att (åter)använda forskningsdata

Vad tror du?

Tack!

Helena Eckerbom
Glenn Haya
Karin Widin

kib@ki.se