

FEGA Sweden

Ett repositoryum för tillgängliggörande av känslig data
från svensk livsvetenskapsforskning

Markus Englund
SNDs nätverksträff i Uppsala 2023-11-07



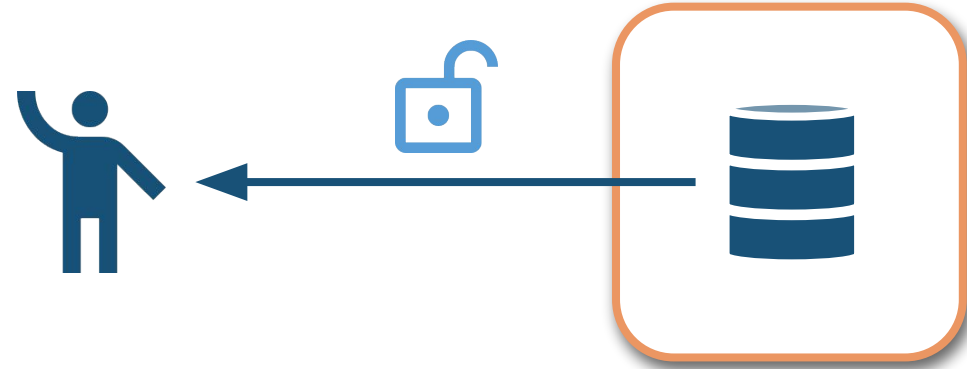
Vad menar vi med “känslig data”?

“känsliga data” handlar i vårt fall om det som man brukar kalla känsliga personuppgifter, det vill säga uppgifter om:

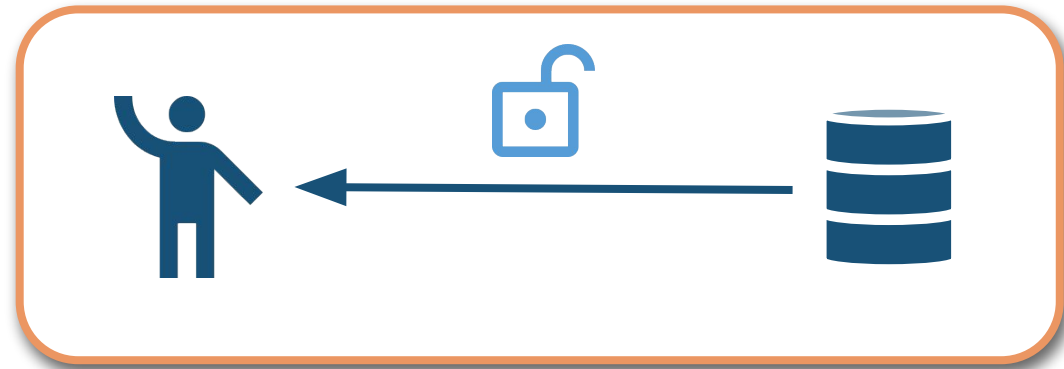
- etniskt ursprung
- politiska åsikter
- religiös eller filosofisk övertygelse
- medlemskap i fackförening
- **uppgifter om hälsa**, sexualliv eller sexuell läggning
- **genetiska uppgifter**
- biometriska uppgifter som entydigt identifierar en person.

Det finns olika sätt att tillgängliggöra känsliga data

Utanför kontrollerad datormiljö



Inom kontrollerad datormiljö



**Vilka problem finns när det gäller att få
och ge tillgång till känsliga
forskningsdata?**

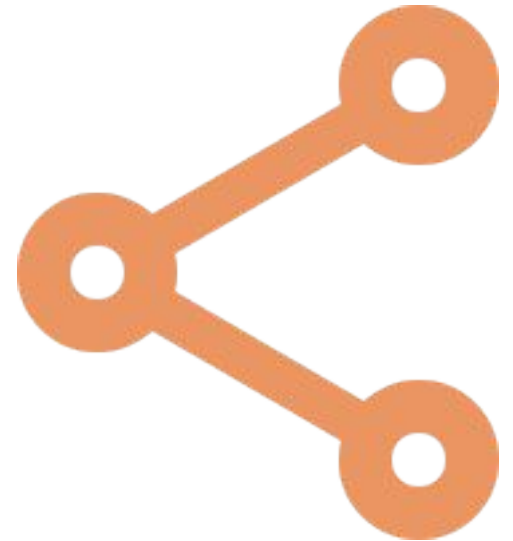
Många hinder för att återanvända känsliga data

- Svårt att hitta dataset
- Oklart hur man går tillväga för att få tillgång till data
- IT-stöd för åtkomst saknas
- Osäkerhet kring vad lagen säger
- Bristande metadata (eftersom data inte deponeras i repositorer)
- Etiskt tillstånd medger inte sekundär användning



Många hinder för att dela känsliga data

- Befintliga repositorer lever inte upp till svenska lärosätens (och andra myndigheters) krav
- Tydliga processer saknas för hur delning ska gå till (vanligt att forskare delar data på egen hand efter projektslut)
- Osäkerhet kring vad lagen säger
- IT-stöd för tillgängliggörande av data saknas
- Plan saknas för hur data ska arkiveras efter projektslut
- Etiskt tillstånd medger inte delning för sekundäranvändning



Hur kan vi underlätta för forskarna?

Vår lösning: FEGA Sweden

- Svensk nod inom Federated European Genome-phenome Archive (FEGA)
- Nationellt domänspecifikt repositorium för att göra data från svensk livsvetenskapsforskning FAIR (Findable, Accessible, Interoperable and Reusable)
- Känsliga data lagras i Sverige (hos Sunet med finansiering från Vetenskapsrådet)
- Inte ett arkiv i arkivlagens mening
- Sjösetts förhoppningsvis under 2023

NBIS svarar för utveckling och användarstöd

- Utveckling av mjukvara sker genom internationellt samarbete
 - Uteslutande öppen källkod
 - Globala standarder (t.ex. GA4GH för kryptering och EGA för metadata)
- Generella biträdesavtal har tecknats med de större lärosätena för att underlätta datadelning
- Helpdesk med data stewards som ger stöd kring deponering och delning av data



Global Alliance
for Genomics & Health

FEGA – ett växande internationellt nätverk

- FEGA är ett formaliserat nätverk av nationella noder och EGA (European Genome-phenome Archive)
- Idag med en handfull nationella noder (bl.a. Finland, Norge, Portugal och Tyskland)
- Samarbete över nationsgränser för att möta forskarnas behov
- Styrs av en strategisk kommitté och en operativ kommitté
- Arbetar aktivt för bättre datahantering, till exempel jobbar man med en ny metadatamodell



FEGA Sweden bidrar till att göra data FAIR

Findable

Dataset tilldelas
accessionsnummer
från EGA

Dataset synliga på
ega-archive.org

Accessible

Publika metadata
på ega-archive.org

System för
accesshantering

Säker överföring av
data

Interoperable

Filformat enligt
rekommendation
från EGA

Publika metadata
enligt EGAs
metadataschema

Reusable

Standardiserade
data och metadata

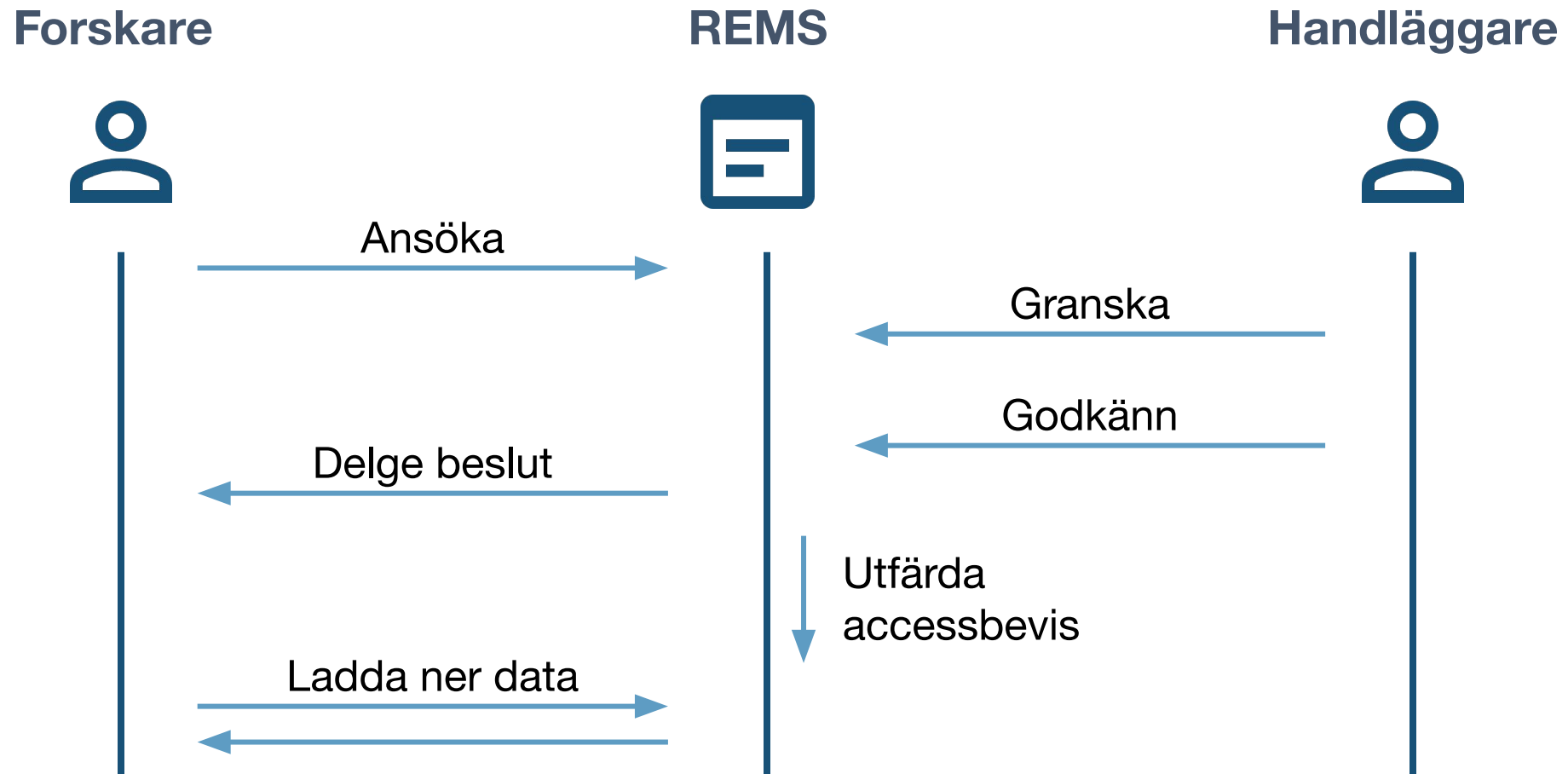
Tydligt vad som
gäller för tillgång
och användning

Hur hanteras access till data i FEGA Sweden?

Accesshantering i FEGA Sweden – så här har vi tänkt

- Access hanteras på dataset-nivå
- Accesshantering sköts av personuppgiftsansvarig (PuA)
 - PuA bestämmer vilka personer som ska ha rätt att bevilja access
 - PuA bestämmer hur processen för beviljande av access ska se ut
- Systemet REMS (Resource Entitlement Management System) används för att hantera access till data:
 - Forskare ansöker om access
 - Handläggare granskar ansökan och bjuder in ytterligare granskare
 - Handläggare beslutar om access
 - Forskaren ges automatiskt tillgång till data när en ansökan beviljats

Accesshantering i FEGA Sweden – så här har vi tänkt



Förslag på ansvarsfördelning vid deponering av data

Forskare

Uttrycker önskemål att deponera data

Registrerar publika metadata

Laddar upp krypterad data

Meddelar när det är dags för release

FEGA Sweden

Tar reda på vem som är personuppgiftsansvarig (PuA)

Informerar DAU (vid PuA) inför deponering

Granskar publika metadata

Informerar DAU inför release

Data Access Unit (DAU)

Registrerar dataset i REMS

Kopplar dataset till handläggare i REMS

Registrerar Data Access Agreement och Data Access Committee i DAC-portalen

Tack!



<https://fega.nbis.se>

Helpdesk: ega-se@nbis.se

Extra bilder



Metadata record

-endast metadata om forskningsmaterialet



SciLifeLab Data Repository

Browse Search the SciLifeLab Data ... Upload My data AA

File(s) not publicly available

Reason: Sequencing was performed at BGI and the sequencing data is stored locally in our lab. We will deposit the WGS data to EGA-SE depository when it is completed.

Single base substitution mutational signatures in pediatric acute myeloid leukemia based on whole genome sequencing

Cite Share + Collect ...

Dataset posted on 04.04.2022, 14:15 authored by [Rebeqa Gunnarsson](#), [Minjun Yang](#), [Linda Olsson Arvidsson](#), [andrea Biloglav](#), [Mikael Behrendtz](#), [Anders Castor](#), [Kajsa Paulsson](#), [Bertil Johansson](#)

USAGE METRICS 159 views 0 downloads

Read the peer review

Single base substitution mutational signatures in pediatric acute myeloid leukemia based on whole genome sequencing

This dataset includes whole genome sequencing (WGS) data of 20 diagnostic, and 20 remission samples from 20 children/adolescents with acute myeloid leukemia (AML), treated at the Departments of Pediatrics at Lund and Linköping University Hospitals between 1994 and 2016. The median age of the patients was 8 years (range 0-17 years) and the female/male ratio was 1:1. DNA was extracted from diagnostic bone marrow (BM; n = 17)/peripheral blood (PB; n = 3) samples, remission BM (n = 15)/PB (n = 5), and from two BM relapses. Construction of libraries, using the TruSeq Nano DNA sample preparation kit (Illumina, San Diego, CA, USA) on 100 ng DNA, and massively parallel

Metadata-records

- Känsliga data
- Data som inte kan delas öppet
- Kriterier för tillgång
- Information om hur man ansöker om tillgång
- Kontaktuppgifter

<https://doi.org/10.17044/scilifelab.13325822.v1>

User case: Datadelningen har haft påverkan



The collage features three news articles and the Swedish Pathogens Portal website. The top-left article, dated 2021-02-16, is titled '96 procent hade kvar corona-antikroppar 9 månader senare' by Niklas Eriksson. The top-right article, dated 2021-02-16, is titled 'Avloppsvattnet avslöjar: Corona skenar i'. The middle-left article, dated 2022-09-07, is titled 'Avloppsprov effektiva för att förutspå pandemi'. The bottom section shows the Swedish Pathogens Portal website, which includes the SciLifeLab logo, navigation menus for 'Available Data & Highlights', 'Data Dashboards', and 'Research & Funding', and a 'Welcome to the new Swedish Pathogens Portal' banner.

- Pandemin har påverkat och förändrat samhället och forskningen.
- Öppen datadelning → nyckel till att mildra pandemins påverkan på samhället.
- Den svenska COVID-19 dataportalen (2020)
 - Nationella noden för **European COVID-19 Data Platform**.
 - Pandemisk beredskap- **SciLifeLab Pandemic Laboratory Preparedness Program**.
- **Swedish Pathogens Portal** (2023)
 - Breddat scope till att omfatta **alla patogener**.
 - pathogens.se/