

Retningslinjer for udstilling af data fra Den Nationale Genomdatabase og regler for hjemtagelse af analyseresultater





Retningslinjer for udstilling af data fra Den Nationale Genomdatabase og regler for hjemtagelse af analyseresultater

Indholdsfortegnelse

1. Introduktion	2
2. Overblik over data i Den Nationale Genomdatabase	2
2.1 Opbygningen af Den Nationale Genomdatabase	2
2.2 Datatyper og dataomfang.....	3
2.3 Datakvalitet.....	3
2.4 Sådan får du et overblik over data i Den Nationale Genomdatabase.....	3
3. Inden du kan få adgang til data fra Den Nationale Genomdatabase	3
3.1 Godkendelse fra det videnskabsetiske komitesystem	3
3.2 Adgang til NGC's infrastruktur	4
3.3 Autorisation	4
3.4 Dataansvar	4
4. Sådan bestiller du data fra Den Nationale Genomdatabase	4
4.1 Ansøgning om udstilling og hjemtagelse af data	4
4.2 Udstilling af data baseret på en CPR-liste	5
4.3 Udstilling af data baseret på metadata.....	5
4.5 Opslag i Vævsanvendelsesregisteret	5
4.6 Projektspecifikt isoleret analysemiljø	5
4.7 Afklaringsmøde	6
4.8 Brugere til NGC's Infrastruktur	6
4.9 Priser	6
5. Sådan arbejder du med data fra Den Nationale Genomdatabase på NGC's infrastruktur	6
6. Sådan kombinerer du data fra Den Nationale Genomdatabase med data fra dit forskningsprojekt	7
7. Sådan hjemtager du analyseresultater baseret på data fra Den Nationale Genomdatabase	7
7.1 Hjemtagelse af de analyseresultater, du ønsker at hjemtage	8
7.2 Kontrol af analyseresultater før hjemtagelse	8
7.3 Hjemtagelse af godkendte analyseresultater	8
8. Regler for hjemtagelse af analyseresultater	9
8.1 CPR-numre	9
8.2 Hvad må hjemtages?	9
8.3 Hvad må <i>ikke</i> hjemtages?	11
9. Tilladelse til at hjemtage data på individniveau	11
10. Sikkerhedskopi	12
11. Procedure og sanktioner ved overtrædelse af NGC's regler for hjemtagelse	12
11.1 Procedurer og sanktioner ved hjemtagelse af individdata	13
11.2 Procedurer og sanktioner ved hjemtagelse af detaljerede tabeller og resultater	13

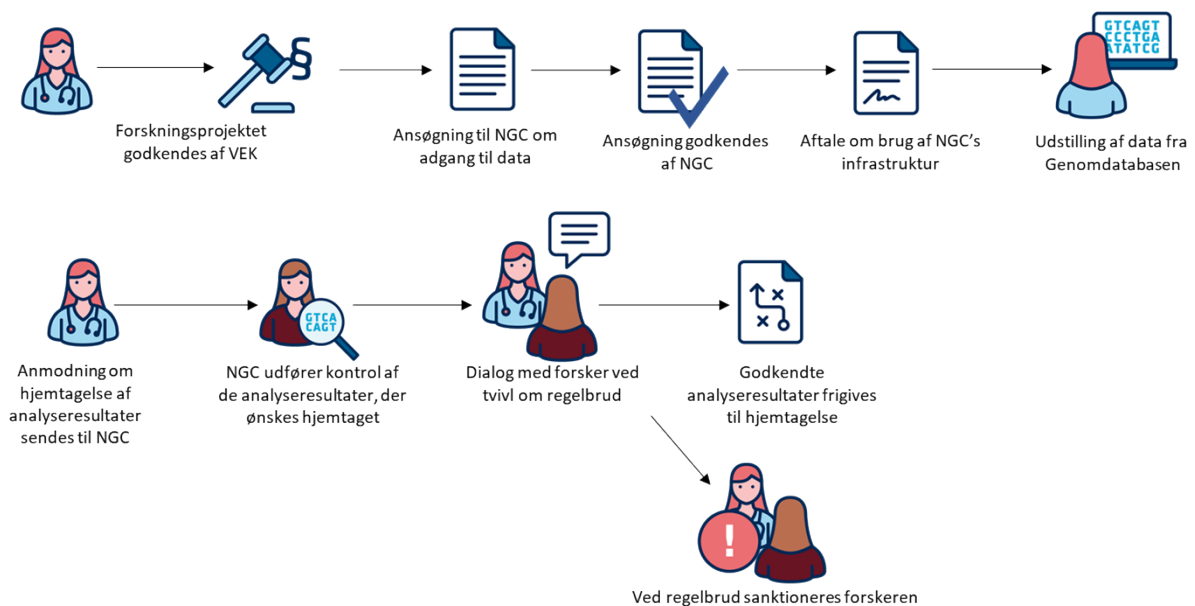
1. Introduktion

Genetiske data fra omfattende genetiske analyser, der er indsamlet i forbindelse med behandling eller forskning i sundhedsvæsenet efter 1. juli 2019, skal indberettes til Nationalt Genom Center og opbevares i Den Nationale Genomdatabase.

Som forsker kan du søge om at få udstillet data fra Den Nationale Genomdatabase til brug for forskning i personlig medicin.

For at få adgang til data fra Den Nationale Genomdatabase skal du igennem en proces med godkendelse af dit projekt hos det videnskabetiske komitesystem, ansøge NGC om adgang til data, arbejde med data på NGC's infrastruktur og til sidst hjemtagelse af analyseresultater til brug for publikationer mm.

Du kan se en oversigt over processen i figur 1. Denne vejledning vil guide dig igennem de enkelte trin.



Figur 1. Oversigt over det administrative forløb fra ansøgning om udstilling af data fra Den Nationale Genomdatabase til hjemtagelse af analyseresultater.

2. Overblik over data i Den Nationale Genomdatabase

Indberetningspligtigt data fra omfattende genetiske analyser skal indberettes til NGC, hvor de opbevares i Den Nationale Genomdatabase. I dette afsnit kan du læse mere om opbygningen af Den Nationale Genomdatabase samt hvilke datatyper, der findes i databasen. Du kan også læse om, hvordan du kan få et overblik over relevant data for dit forskningsprojekt.

2.1 Opbygningen af Den Nationale Genomdatabase

Den Nationale Genomdatabase er placeret på NGC's Infrastruktur og består teknisk af en object store (S3) til rådata (fx FASTQ-filer) og en relationel database (PostgreSQL) til metadata. Dette design tilbyder effektiv søgbarhed i genomdatabasen via metadata.

2.2 Datatyper og dataomfang

Den Nationale Genomdatabase indeholder data på sekventeret prøve-niveau. Dette omfatter rådata fra sekventering samt tilhørende metadata. Der tilføjes på nuværende tidspunkt ikke yderligere metadata fra eksterne systemer eller registre, og det er derfor et begrænset sæt metadata, som stilles til rådighed for brugere. NGC vedligeholder en oversigt over aktuelle metadata, som modtages via indberetningen for hver sekventeret prøve. Genomdatabase indeholder rådata og tilhørende metadata fra NNF WGS og Total RNA-sekventeringer i et klinisk flow. Rådata indberettes specifikt som FASTQ-filer (i "read pairs") fra Illumina sekventeringsmaskiner. Genomdatabase indeholder ikke analyserede data (fx output fra bioinformatiske pipelines), herunder QC-parametre eller information om varianter.

2.3 Datakvalitet

NGC foretager en række tjek i indberetningsflowet. Fx valideres det, at de krævede filer er indeholdt i en given indberetning, at kobling mellem rådata og metadata er korrekt, samt at metadata overholder specifikationerne (fx formater, krævede og valgfri felter m.v.). NGC vedligeholder en oversigt over gældende metadata-specifikationer. Det kan ikke garanteres, at alle fejl identificeres, og det er indberettets ansvar, at alle data er korrekte ved indberetning. NGC gemmer indberettede data i genomdatabase under denne forudsætning. NGC modtager ikke information om kvaliteten af sekventeringer. Fx kendes ikke sekventeringsdybden for en given WGS.

2.4 Sådan får du et overblik over data i Den Nationale Genomdatabase

Inden du går i gang med at søge det videnskabetiske komitesystem om godkendelse af adgang til data fra Den Nationale Genomdatabase, har du mulighed for at få et overblik over, om databasen indeholder data relevant for dit forskningsprojekt.

Du kan få et overblik over data inden for nogle metadata-kriterier, fx antal genomer inden for en bestemt patientgruppe. Du kan læse mere om de metadata, der er tilgængelige i Den Nationale Genomdatabase her: [Udstilling af data baseret på metadata](#)

Hvis du ønsker et overblik over, om Den Nationale Database indeholder data relevant for dit forskningsprojekt, skal du sende en mail med dine ønsker til data til forskning@ngc.dk.

3. Inden du kan få adgang til data fra Den Nationale Genomdatabase

For at kunne få udstillet data fra Den Nationale Genomdatabase skal en række kriterier være opfyldt. Disse er beskrevet i dette afsnit.

3.1 Godkendelse fra det videnskabetiske komitesystem

For at få udstillet data fra Den Nationale Genomdatabase, skal dit forskningsprojekt være godkendt af det videnskabetiske komitesystem. Hvilken komite, der skal godkende afhænger af projektets udformning. I din protokol skal du anføre, præcis hvilke data fra Den Nationale Genomdatabase, du ønsker adgang til.

Særligt for hjemtagelse af personhenførbare data

I udgangspunktet må du ikke hjemtage personhenførbare data fra Den Nationale Genomdatabase. Skulle dit forskningsprojekt have særligt behov for dette, kræver det godkendelse fra det videnskabetiske komitesystem. Hvis du derfor allerede inden projektstart ved, at du vil hjemtage personhenførbare data, kan du med fordel skrive det ind i din protokol fra start. Du kan læse mere om, hvad der kræves for at hjemtage personhenførbare data her: [Tilladelse til at hjemtage data på individniveau](#).

3.2 Adgang til NGC's infrastruktur

For at få udstillet data fra Den Nationale Genomdatabase skal dit forskningsprojekt have en aftale med NGC om adgang til en cloud på NGC's Infrastruktur.

Hvis du ikke allerede har en aftale med NGC, kan du læse mere om at få adgang til NGC's Infrastruktur samt ansøgningsproceduren her: [Søg om adgang \(ngc.dk\)](#)

For at få adgang til en cloud skal du forske i personlig medicin og udføre videnskabelige undersøgelser af væsentlig samfundsmæssig betydning i samarbejde med en dansk forskningsinstitution.

Hvis du allerede har en forudgående aftale med NGC, skal du i stedet gå direkte til [Bestilling af data, der ønskes udstillet](#)

3.3 Autorisation

NGC kan som udgangspunkt kun give adgang til NGC's Infrastruktur til forskningsprojekter, der er tilknyttet en autoriseret dansk forskningsinstitution.

Institutionen kan være autoriseret hos enten hos Danmarks Statistik eller Sundhedsdatastyrelsen.

Hvis den institution, forskningsprojektet er tilknyttet, ikke har en autorisation hos en af disse, skal institutionen autoriseres hos NGC.

Bemærk, at ved at indsende oplysninger om allerede eksisterende autorisation, accepterer du, at NGC må udveksle oplysningerne med den myndighed, hvor dit miljø er autoriseret.

De autoriserede institutioner er ansvarlige for, at de brugere, der arbejder under autorisationen, har de nødvendige kompetencer, og at de er godt bekendt med NGC's regler for hjemtagelse af analyseresultater baseret på data fra Den Nationale Genomdatabase.

3.4 Dataansvar

NGC er dataansvarlig for det data, der ligger i Den Nationale Genomdatabase.

Når et forskningsprojekt får udstillet data fra Den Nationale Genomdatabase, bliver projektet dataansvarlig for det data, der udstilles, og NGC bliver databehandler.

Der sker altså en videregivelse af data. En databehandleraftale regulerer dette forhold. Hvis du allerede har en aftale med NGC om adgang til NGC's Infrastruktur, er en relevant databehandleraftale en del af dette aftalegrundlag.

4. Sådan bestiller du data fra Den Nationale Genomdatabase

4.1 Ansøgning om udstilling og hjemtagelse af data

Du kan få udstillet data fra Den Nationale Genomdatabase i overensstemmelse med din godkendte protokol til det videnskabetiske komitesystem. Du skal sende et udfyldt ansøgningsskema til forskning@ngc.dk. I ansøgningen skal du beskrive, hvilke data du ønsker udstillet, og hvilke slags data du vil hjemtage.

Ansøgningsskemaet vedr. udstilling af data fra Den Nationale Genomdatabase kan

findes her: [Ansøgning om udstilling af data fra Den Nationale Genomdatabase \(ngc.dk\)](#). Her kan du også læse mere om ansøgningsprocessen.

4.2 Udstilling af data baseret på en CPR-liste

Du kan få udstillet data fra Den Nationale Genomdatabase baseret på en CPR-liste, dvs. du har en liste med CPR-numre, hvor du ønsker data udstillet fra netop disse personer. Listen med CPR-numre sendes med sikker mail til NGC Forskerservice på forskning@ngc.dk. Hvis du ikke har mulighed for at sende med sikker mail kan CPR-listen i stedet sendes krypteret med kode, hvorefter du aftaler med NGC Forskerservice, hvordan koden til filen sendes.

CPR-listen skal sendes som en (semikolon-separeret) csv-fil, og der er følgende krav til filen:

- Ingen mellemrum i filnavn
- Linje 1 skal indeholde projektakronym uden mellemrum
- Linje 2 skal indeholde kolonneoverskrifter (serienummer, CPR-nummer)
- Kolonne A skal indeholde indekseringsnumre (fx 1, 2, 3...)
- Kolonne B skal indeholde CPR-nummer (med eller uden bindestreg)
- Ingen eCPR

Hvis du har nogle spørgsmål til opbygningen af CPR-listen, kan du kontakte NGC Forskerservice på forskning@ngc.dk

4.3 Udstilling af data baseret på metadata

Du kan få udstillet data fra Den Nationale Genomdatabase ud fra nogle metadata-kriterier. Det kan fx være, at du ønsker data fra patienter med en bestemt alder eller køn tilhørende en bestemt patientgruppe.

Hvilke metadata du ønsker data baseret på, skal du beskrive i din ansøgning om udstilling af data fra Den Nationale Genomdatabase.

De metadata, som du kan få udstillet data på baggrund af, er bl.a.:

- Alder
- Køn
- Patientgruppe (se de 17 patientgrupper her: [Patientgrupper \(ngc.dk\)](#))
- Prøvetype (germline eller somatisk)
- Materialetype (fx blod eller væv)
- Sekventeringsdato

Den aktuelle liste over tilgængeligt metadata kan rekvireres af NGC ved at skrive til forskning@ngc.dk.

4.5 Opslag i Vævsanvendelsesregisteret

Alle CPR-numre bliver sammenkørt med Vævsanvendelsesregisteret (VAR) før udstilling. Hvis en person er registeret i VAR, kan du ikke få udstillet data fra denne person.

Du vil få besked om, hvor mange CPR-numre, der er filteret fra grundet registrering i VAR, og hvor mange CPR-numre fra din CPR-liste eller fra dine metadata-kriterier, der er fundet data på i Den Nationale Genomdatabase. På baggrund af dette vil du blive bedt om at vurdere, om du stadig ønsker data udstillet på NGC Infrastruktur.

4.6 Projektspecifikt isoleret analysemiljø

NGC er dataansvarlig for data, der opbevares i Den Nationale Genomdatabase.

Samtidig er det pålagt NGC at udvikle og drive en fælles, landsdækkende informationsinfrastruktur, hvor der ikke vil ske en fysisk udlevering af data fra Den Nationale Genomdatabase. Derfor bliver data fra Den Nationale Genomdatabase

udstillet i et projektspecifikt isoleret analysemiljø til dit forskningsprojekt. Det betyder, at selvom du allerede har en NGC Cloud, vil du få oprettet en ny cloud til det udstillede data.

Det isolerede analysemiljø er opbygget som et secure processing environment, hvor der ikke er adgang til internettet, og hvorfra data fra genomdatabaseen ikke kan downloades eller eksporteres.

Hvis du skal eksportere noget fra dit isolerede analysemiljø skal det hjemtages via NGC's hjemtagelsesprocedure, som du kan læse mere om her: [Hjemtagelse af analyseresultater](#).

4.7 Afklaringsmøde

Når du har fået tilsagn fra NGC om adgang til data fra Den Nationale Genomdatabase, vil du blive indkaldt til et møde med NGC. På dette møde vil vi forklare mere om, hvordan du får stillet data til rådighed samt afklare en række ting bl.a. hvilke værktøjer, du har brug for, om du ønsker at kombinere data fra Den Nationale Genomdatabase med data genereret i projektet, om du har brug for bioinformatisk konsulentbistand mm.

På dette møde vil du kunne sparre med NGC's rådgivere, så vi sikrer den optimale løsning for jeres forskningsprojekt.

4.8 Brugere til NGC's Infrastruktur

Hvis du ønsker adgang til det udstillede data, skal du have en bruger til NGC's HPC. Din bruger skal have særlig adgang til dit projektspecifikke isolerede analysemiljø, hvor data til dit forskningsprojekt udstilles. Det er forskningsprojektets PI, der giver tilladelse til, hvem der har adgang til udstillet data fra Den Nationale Genomdatabase i projektets isolerede analysemiljø.

Det er den projektansvarliges ansvar, at alle brugere med adgang til udstillet data fra Den Nationale Genomdatabase er bekendt med NGC's retningslinjer for at arbejde med udstillet data og regler for hjemtagelse af analyseresultater.

Hvis du ikke allerede har en bruger til NGC's HPC skal du sende en udfyldt brugerblanket ([Blanketter og vejledninger \(ngc.dk\)](#)) med den projektansvarliges underskrift til forskning@ngc.dk.

Hvis du allerede har en bruger, skal du sende en mail til forskning@ngc.dk, hvor du anmoder om adgang til projektets isolerede analysemiljø med godkendelse fra projektansvarlig.

4.9 Priser

Det er gratis at få udstillet data fra Den Nationale Genomdatabase. Men du skal betale for storage og compute-forbrug i dit projektspecifikke isolerede analysemiljø, jf. NGC's prismodel. Forbrug i et isoleret analysemiljø faktureres som en NGC-managed Cloud. Prismodellen kan ses her: [NGC forskerservice](#)

Hvis du ikke allerede har en aftale om brug af NGC's Infrastruktur, er der et administrationsgebyr på 5.000 kr. for indgåelse af aftale.

5. Sådan arbejder du med data fra Den Nationale Genomdatabase på NGC's infrastruktur

Når NGC har identificeret de relevante data til dit forskningsprojekt, bliver data udstillet til dig i dit projektspecifikke isolerede analysemiljø på NGC's Infrastruktur.

I dit isolerede analysemiljø har du adgang til de open source-værktøjer, som er

tilgængelige på NGC's Infrastruktur. Det isolerede analysemiljø er et lukket miljø uden adgang til internettet.

Det er kun tilladt at arbejde med data fra Den Nationale Genomdatabase på arbejdspladsen eller via en hjemmearbejdsplads.

I udgangspunktet udstilles data fra Den Nationale Genomdatabase i pseudonymiseret form. Det er tilladt både at se og arbejde med pseudonymiseret individdata på skærmen. Men du må aldrig forsøge at identificere enkeltpersoner - heller ikke dig selv.

Hvis du har brug for hjælp til programmering, er det tilladt at få hjælp af en anden person, der også har godkendt adgang til data. Du må dog aldrig forlade skærmen, så længe hjælpen pågår.

6. Sådan kombinerer du data fra Den Nationale Genomdatabase med data fra dit forskningsprojekt

Hvis du ønsker at kombinere data fra Den Nationale Genomdatabase med data, du selv har indsamlet i forbindelse med dit forskningsprojekt (fx eget indsamlet genomdata eller registerdata), skal dette gøres i dit projektspecifikke isolerede analysemiljø.

Data, der placeres i det isolerede analysemiljø, kan ikke eksporteres fra mappen igen uden at gå igennem NGC's hjemtagelsesprocedure. Du bør derfor overveje, hvilke data, du vil have placeret i det isolerede analysemiljø, og om det vil være relevant at beholde en kopi af dit eget hos dig selv eller på dit forskningsprojekts NGC Cloud.

Det er NGC, der står for kopieringen af data til det isolerede analysemiljø under instruks fra forskningsprojektets PI. Flytningen af data foregår via SFTP.

Hvis du ønsker at placere data i dit isolerede analysemiljø, bedes du sende en mail til forskning@ngc.dk, hvorefter vi aftaler processen for kopiering af data til dit isolerede analysemiljø.

Hvis dit forskningsprojekt har sin egen cloud på NGC's Infrastruktur vil der være nogle ekstra udgifter til forskningsprojektet, når I placerer en kopi af data i det isolerede analysemiljø til udstilling af data. For at minimere udgifterne til storage anbefaler vi følgende:

- Original-data som ligger i dit forskningsprojekts egen cloud kan placeres i cold storage (archive)
- Slette kopien af data i det isolerede analysemiljø, så snart du ikke skal bruge det mere.

7. Sådan hjemtager du analyseresultater baseret på data fra Den Nationale Genomdatabase

Når du har arbejdet med det udstillede data fra Den Nationale Genomdatabase og har generet nogle analyseresultater, som du ønsker at hjemtage fra NGC's Infrastruktur, fx til publicering, skal du følge NGC's regler for hjemtagelse af analyseresultater. Alt data, der ønskes hjemtaget, skal foregå via NGC's hjemtagelsesprocedure (figur 2, s. 9). Dvs. det er ikke tilladt at skrive oplysninger fra skærmen, tage et skærmprent eller et billede af skærmen med mobiltelefonen.

Er du i tvivl om, hvorvidt du må hjemtage en fil, kan du kontakte NGC Forskerservice på forskning@ngc.dk

7.1 Hjemtagelse af de analyseresultater, du ønsker at hjemtage

Når du har analyseresultater klar til hjemtagelse, skal du sende en mail til forsknings@ngc.dk, med besked om, at du har analyseresultater klar til hjemtagelse.

Med mailen skal du også sende:

- NGC's hjemtagelsesformular med underskrift fra projektets PI og dig som hjemtager analyseresultater, som forsikrer, at I begge er bekendt med NGC's [Regler for hjemtagelse af analyseresultater](#) samt [Procedure og sanktioner ved overtrædelse af NGC's regler for hjemtagelse](#)
- VEK-godkendelse af dit forskningsprojekt
- Evt. samtykke fra patienter eller dispensation for samtykke, ved hjemtagelse af personhenførbare data
- Manifest over analyseresultater, der ønskes hjemtaget. Manifestet skal indeholde:
 - Præcis stinavn for alle analyseresultater
 - Beskrivelse af hvert analyseresultat, samt en beskrivelse af, hvorfor de overholder NGC's regler for hjemtagelse (fx alt data er aggregeret, der er intet individdata)
 - Hvis relevant, en beskrivelse af hvilke analyseresultater, der indeholder personhenførbare data

NGC's hjemtagelsesformular, finder du her: [Formular til hjemtagelse af analyseresultater baseret på data fra Den Nationale Genomdatabase \(ngc.dk\)](#).

Template til manifest kan rekvireres hos forskning@ngc.dk.

NGC vil flytte de ønskede analyseresultater fra dit isoleret analysemiljø til NGC's slusemappe, hvor du ikke længere har adgang til analyseresultaterne (figur 2, s. 9). I denne slusemappe har NGC Forskerservice læseadgang til analyseresultaterne for at kunne udføre kontrol af, at kun tilladte analyseresultater hjemtages.

7.2 Kontrol af analyseresultater før hjemtagelse

Baseret på den fremsendte formular og manifestet kontrollerer NGC Forskerservice, at du har de nødvendige tilladelser til at må hjemtage de ønskede analyseresultater samt kontrollerer, at analyseresultaterne stemmer overens med beskrivelserne i formularen og overholder NGC's [Regler for hjemtagelse af analyseresultater](#).

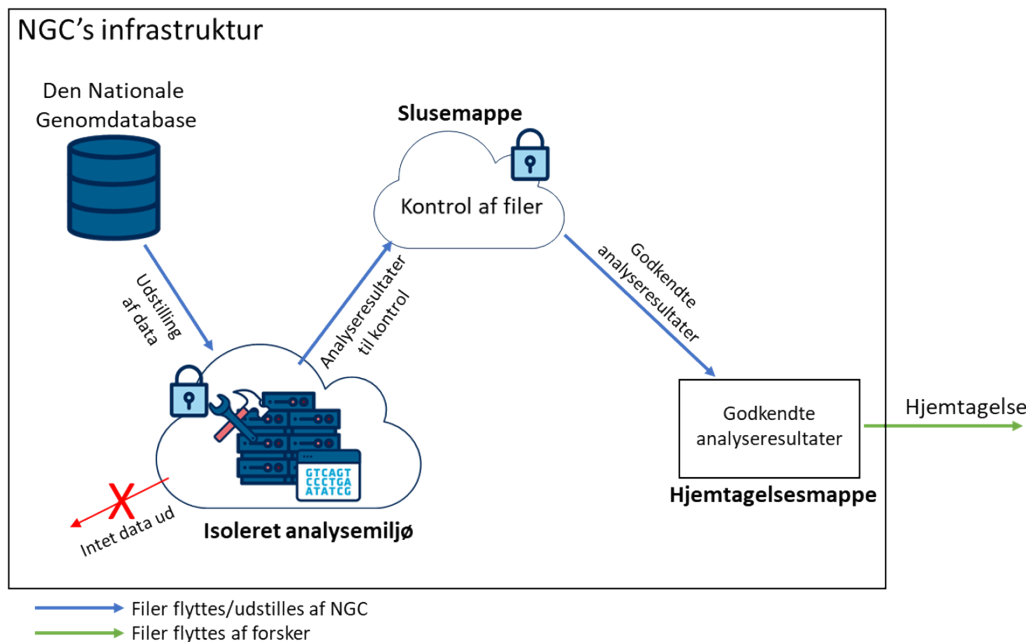
Behandlingstiden for kontrol af data er maksimalt 10 arbejdsdage, men forventeligt kortere.

Ved tvivl om, hvorvidt dataindholdet i analyseresultaterne overholder NGC's regler for hjemtagelser, eller ved tvivl om, hvorvidt analyseresultaterne indeholder (ikke-tilladte) personhenførbare data, kan NGC Forskerservice tage kontakt til dig for en uddybende forklaring af dataindholdet i analyseresultaterne.

7.3 Hjemtagelse af godkendte analyseresultater

De analyseresultater, som NGC godkender til hjemtagelse, bliver placeret i en hjemtagelsesmappe oprettet til dit forskningsprojekt. Hjemtagelsesmappen er en SFTP-mappe, hvorfra du har mulighed for at hjemtage analyseresultater fra NGC's Infrastruktur (figur 2, s. 9). Du modtager en mail fra NGC Forskerservice med besked om, at de godkendte analyseresultater ligger klar til hjemtagelse.

Analyseresultaterne ligger i hjemtagelsesmappen i 30 dage, inden de slettes. Du skal derfor huske at hjemtage dine analyseresultater indenfor 30 dage.



Figur 2. Oversigt over dataflowet i NGC's hjemtagelsesprocedure.

8. Regler for hjemtagelse af analyseresultater

NGC er dataansvarlig for data, der opbevares i Den Nationale Genomdatabase. Samtidig er det pålagt NGC at udvikle og drive en fælles, landsdækkende informationsinfrastruktur, hvor der ikke vil ske en fysisk udlevering af data fra Den Nationale Genomdatabase. Hos NGC bliver der således arbejdet udelukkende med data i dertil dedikerede isolerede analysemiljøer, hvorfra data fra genomdatabasen ikke kan downloades eller eksporteres. Videre har den enkelte borger krav på, at sundhedsdata behandles med den største omhyggelighed. Derfor har NGC opsat en række regler for, hvad der må hjemtages fra infrastrukturen, hvis de analyseresultater, der ønskes hjemtaget, er baseret på data fra Den Nationale Genomdatabase. Reglerne er beskrevet nedenfor.

Overtrædelser af reglerne kan betyde lukning af data- og brugeradgange for dig og alle andre brugere på dit projekts cloud og potentielt for alle brugere med adgang til data på NGC's HPC under samme autoriserede institution.

8.1 CPR-numre

Data fra Den Nationale Genomdatabase udstilles som udgangspunkt i pseudonymiseret form. Du får derfor ikke adgang til CPR-numre.

Hvis du har brug for CPR-numre, skal du have godkendelse fra det videnskabetiske komitéssystem til at må få udleveret CPR-numre.

Ligeledes skal du argumentere for dette i din ansøgning til NGC om adgang til data fra Den Nationale Genomdatabase. Du kan kun få lov at hjemtage CPR-numre, hvis det er absolut nødvendigt.

8.2 Hvad må hjemtages?

Som hovedregel skal alle analyseresultater baseret på data fra Den Nationale Genomdatabase være aggregerede i en sådan grad, at det ikke er muligt at identificere enkelte enheder, dvs. hverken personer, familier eller andre enheder. Generelt gælder nedenstående tommelfingerregler for aggregerede data:

Hjemsendte analyseresultater bør som udgangspunkt have sådan en form, at de umiddelbart kan anvendes i en publikation.

Tabeller:

Tabeller bør indeholde mindst 5 observationer pr. celle. For tabeller er tommelfingerreglen, at tabeller skal indeholde mindst 5 observationer, men aggregeringsniveauet skal altid afhænge af en konkret vurdering af, hvad tabellen indeholder. Max. og min. og percentiler, som fx median, må kun hjemsendes, hvis der ikke er fare for identifikation af enkeltindivider. Ingen personer må kunne genkendes i det hjemsendte output.

Figurer:

Figurer må kun hjemsendes, hvis de ikke indeholder information, som kan identificere enkeltindivider. Man skal ved figurer være særligt opmærksomme på ekstreme værdier, såkaldte outliers. Det skal derudover sikres, at figurerne ikke indeholder indlejrede mikrodata.

Filtyper:

Rå-data må ikke hjemtages fra NGC's Infrastruktur. Dette gælder FASTQ-filer og bl.a. VCF- og CRAM-filer.

Tabel 1 viser en liste over filtyper, det er tilladt at hjemtage. Hvis du ønsker at få en ny filtype tilføjet til listen, skal du kontakte forskning@ngc.dk

Filtype	Fil-ekstension
Excel	xls xlsx xlsx xlm xml
Grafik	png tif jpg gif jpeg svg bmp ppt pptx pptm tiff
HTML	htm html
Programkode	R py
Logs	Log
Tekst	Txt tab
PDF	Pdf
Word	doc docx rtf

Tabel 1. Liste over filtyper, det er tilladt at hjemtage fra NGC's Infrastruktur

8.3 Hvad må *ikke* hjemtages?

Ingen form for individdata må hjemtages.

De analyseresultater, der generes baseret på data fra Den Nationale Genomdatabase må i udgangspunktet ikke hjemtages, hvis det er muligt at identificere enkeltindivider.

Ingen form for individdata må hjemtages, hverken via NGC's hjemtagelsesprocedure eller på andre måder fx ved at skrive oplysninger fra skærmen, tage et skærmpoint eller et billede af skærmen med mobiltelefonen.

Vær opmærksom på programmer og logfiler

Programmer og logfiler kan også indeholde individdata fx:

- Programkode der indeholder individdata skrevet ind i en betingelse
- Logfiler der lister observationer. Nogle procedurer lister fx et lille udsnit af det datasæt, der er anvendt til analysen

Hvis man sender en logfil hjem, der lister pseudonymiserede personnumre er det et brud på reglerne – også selv om disse pseudonymiserede personnumre ikke indeholder anden information.

Ingen ID'er der er tilknyttet et enkelt individ må hjemtages

Da analyseresultater ikke må indeholde individdata, er det ikke tilladt, at resultaterne indeholder nogen form for ID'er, der kan identificere en person. Det kan fx være CPR-numre, patientID eller deltagerID.

Detaljerede tabeller og resultater bør ikke hjemtages

Tabeller og resultater der indeholder information, der gør det muligt at identificere enkelt individer, må ikke hjemtages. Som udgangspunkt er det derfor ikke tilladt at hjemtage tabeller og resultater, der indeholder celler med mindre end fem observationer.

Indeholder et eller flere felter i en tabel færre end fem observationer, skal du overveje at aggregere disse tabeller yderligere, alternativt blænde celler med få observationer før tabellen hjemsendes. Afblanding kan fx ske ved at slette cellens indhold eller erstatte det med en markering såsom "n<5"

Vær opmærksom på, at man heller ikke af en tabels marginale totaler, procenter eller rater må kunne udlede, om en given information stammer fra for få observationer. Fx må der ikke af en tabels summer kunne udledes hvor mange personer, der er blevet diskretioneret. Man må med andre ord ikke kunne "regne tilbage".

Modsat individdata er det, når det gælder tabeller og resultater, ikke muligt at lave helt faste regler for, hvad der må hjemtages, og hvad der ikke må. Det er med andre ord tilladt at hjemtage tabeller, figurer mv., der har en større detaljeringsgrad end de tommelfingerregler, som vi anbefaler, hvis det er sikret, at det ikke er muligt at identificere enkeltpersoner i de hjemtagne filer. Fx er det tilladt at hjemtage medianen, selv om det reelt er mikrodata, hvis det er sikret, at ingen personer kan genkendes i materialet. Det er dog vigtigt at understrege, at al hjemtagelse sker på eget ansvar.

9. Tilladelse til at hjemtage data på individniveau

Ved forskning i genomdata vil man ikke altid kunne overholde kravet om, at individdata ikke må hjemtages fra NGC's Infrastruktur. Forskere kan have behov for

at hjemtage data, hvor det faktisk er muligt at identificere enkeltindivider (fx hvis der forskes i sjældne varianter). I disse tilfælde vil analyseresultater, der hjemtages, ikke kunne pseudonymiseres.

Hvis du ønsker at hjemtage analyseresultater, hvor det vil være muligt at identificere enkeltindivider, skal du:

- Have beskrevet i din forskningsprotokol, at du vil hjemtage analyseresultater baseret på data fra Den Nationale Genomdatabase fra NGC's Infrastruktur, som indeholder personhenførbare data.
- Have samtykke fra patienten til at må hjemtage personhenførbare data, hvor der er risiko for, at patienten kan identificeres. Eller have fået dispensation for samtykke af det videnskabetiske komitesystem.
- Beskrive argumenterne for behovet for at hjemtage individdata i din hjemtagelsesformular.
- Hvis du har behov for CPR-numre, skal dette også tydeligt beskrives i din protokol og i hjemtagelsesformularen.

NGC Forskerservice vurderer, om du har berettigelse til at hjemtage analyseresultater baseret på data fra Den Nationale Genomdatabase, der indeholder personhenførbare data.

Hvis du får godkendelse til at må hjemtage personhenførbare data, kontrolleres dine analyseresultater til hjemtagelse på baggrund af denne godkendelse. Du har ansvaret for, at personhenførbare data opbevares sikkert efter hjemtagelse fra NGC's Infrastruktur.

10. Sikkerhedskopi

NGC tager sikkerhedskopi af dit isolerede analysemiljø i henhold til NGC's procedure. Al data bliver replikeret til backup off-site. Denne replikering udføres i form af snapshots, der tages flere gange dagligt. Disse snapshots har en levetid på 30 dage, hvorefter de endeligt slettes. Det er derfor muligt at gendanne on-site data, der er blevet slettet eller ændret, indenfor et vindue på **30 dage**, hvorefter data vil være **endeligt slettet**.

11. Procedure og sanktioner ved overtrædelse af NGC's regler for hjemtagelse

NGC placerer alle analyseresultater, som du anmoder om at hjemtage, i slusemappen, hvor de kontrolleres af NGC. Først når analyseresultaterne er kontrollerede og godkendt af NGC, placeres de i hjemtagelsesmappen, hvorfra du kan hjemtage analyseresultater.

Når du anmoder om hjemtagelse af analyseresultater, har du ansvaret for, at analyseresultaterne overholder NGC's [Regler for hjemtagelse af analyseresultater](#).

Hvis du mener, at reglerne kan være overtrådt, skal du hurtigst muligt kontakte NGC på forskning@ngc.dk og oplyse, hvilke analyseresultater det drejer sig om.

Ved brud på NGC's regler for hjemtagelse kan sanktioner pålægges dig som bruger og det forskningsprojekt, du er tilknyttet. Vi skelner i den forbindelse mellem hjemtagelse af individdata og hjemtagelse af detaljerede tabeller og resultater.

11.1 Procedurer og sanktioner ved hjemtagelse af individdata

Hvis vi konstaterer, at der er anmodet om hjemtagelse af individdata (fx ID'er eller data på individniveau), som projektet ikke har tilladelse til at hjemtage, lukker vi omgående for brugerens adgang til NGC's HPC. Derudover bliver der lukket for adgang til forskningsprojektets isolerede analysemiljø for alle brugere på miljøet. Ved gentagne regelbrud vil yderste konsekvens være, at hele institutionen, som forskningsprojektet er autoriseret under, bliver udelukket for adgang til NGC's HPC.

Forskeren, der har anmodet om hjemtagelse, samt projektets projektansvarlig vil blive kontaktet og blive oplyst om, at alle analyseresultater, der overtræder reglerne for hjemsendelse af analyseresultater, er blevet flyttet fra slusemappen tilbage i projektets isolerede analysemiljø.

NGC vil derudover udbede sig en redegørelse for regelbruddet samt omfanget af anmodede analyseresultater, der bryder reglerne. Det er projektansvarlig på projektet, der har ansvaret for, at redegørelsen bliver udarbejdet. Efterfølgende vil NGC bede projektansvarlig om snarest muligt at sende en plan for, hvordan projektet fremadrettet vil undgå yderligere regelbrud.

Alle adgange forbliver lukket, indtil både redegørelsen og planen er modtaget. Herefter beslutter NGC's direktion, hvilken yderligere sanktion projektet skal pålægges. Sagsbehandlingen kan tage op til 5 arbejdsdage.

En vejledende oversigt over de mulige sanktioner kan ses i tabel 2 på side 14.

Hvis brugeren selv gør NGC opmærksom på, at vedkommende har anmodet om hjemtagelse af analyseresultater, der ikke overholder NGC's regler for hjemtagelse, vil det blive betragtet som en formildende omstændighed, og det vil muliggøre, at adgangen til NGC's HPC vil kunne åbnes igen, så snart NGC fået klarhed over overtrædelsens omfang og karakter.

11.2 Procedurer og sanktioner ved hjemtagelse af detaljerede tabeller og resultater

Hvis NGC konstaterer, at der er anmodet om hjemtagelse af tabeller eller figurer, hvor der er risiko for identificering af individet, lukkes der IKKE for adgangen. NGC vil kontakte brugeren for en redegørelse og vejlede brugeren i reglerne for hjemtagelse af analyseresultater.

Hvis der er tale om gentagne brud på aggregerings-kravet, eller hvis detaljeringsgraden i analyseresultater er meget høj (fx hvis langt de fleste celler indeholder 1-taller), så kan sagen behandles som hjemtagelse af individdata.

Ved enkeltstående overtrædelser vil det for brugere og institutionen normalt ikke få konsekvenser for adgangen til NGC's HPC, men ved gentagne regelbrud, hvor der anmodes om hjemtagelse af analyseresultater, hvor der er fare for identifikation af enkeltpersoner, kan denne praksis føre til lukning af projektets adgang til deres isolerede analysemiljø og en videre behandling af sagen svarende til hjemsendelse af individdata.

	Sanktion mod bruger		Sanktion mod projekt		
Sanktion	Brugerens forsker adgang til NGC's HPC lukkes		Alle projektets brugeres forsker adgang på NGC's HPC lukkes		
	1. gang	Flere gange	1. gang	2. gang	Gentagne gange
Hjemsendelse af individdata					
a. Uheld eller ubetænksom handling	Frem til redegørelse kan godkendes*	1 måneds karantæne	Frem til redegørelse kan godkendes*	1 måneds karantæne	3 måneders karantæne
b. Bevidst handling, fx bevidst forsøg på at identificere individer	3 måneders karantæne	3 måneders karantæne Mulighed for permanent udelukkelse	3 måneders karantæne	6 måneders karantæne Mulighed for permanent udelukkelse	Permanent udelukkelse
Password eller adgang videregivet af autoriseret person					
a. Med vilje	3 måneders karantæne	Permanent udelukkelse	1 måneds karantæne	3 måneders karantæne	Permanent udelukkelse

*Det betragtes som en formildende omstændighed, hvis projektet eller brugeren selv underretter om fejlen

Tabel 2. Vejledende oversigt over sanktioner ved regelbrud.