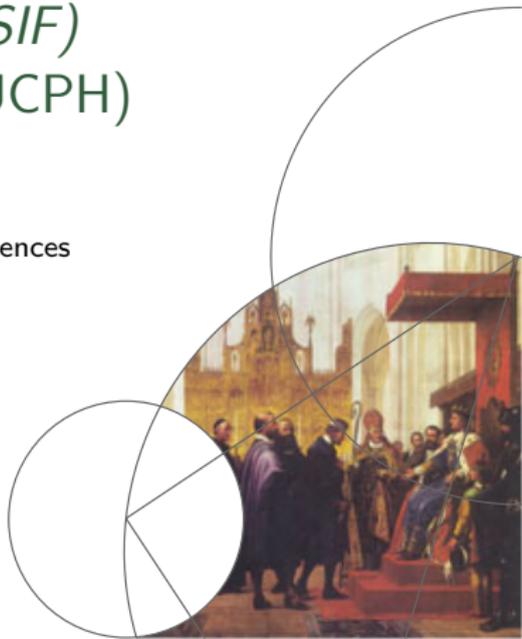




Personal experiences with *Sensitive Information Facility (SIF)* at University of Copenhagen (UCPH)

Anders Tolver, Associate Professor
Data Science Lab, Department of Mathematical Sciences



My needs for storing sensitive data

- I am involved in research projects with personal data
- I process personal data for researchers employed at Capital Region (RegionH)
- I need to collaborate on the analysis of data and to share results with
 - researchers at Capital Region
 - student assistants employed locally at our department

The Research Portal at KUnet lists five ways to store data in accordance with requirements from GDPR including:

SIF (Sensitive Information Facility)

(local KU-version of DeiC Sensitive Storage)



https://sif.ku.dk/



Sensitive Information Facility
University of Copenhagen



English

Welcome to SIF

KU / UCPH Users External Users

Sign up to SIF using your KU / UCPH account?
[sign up](#)

I'm already signed up to SIF with my KU / UCPH account!
[log in](#)

About SIF

SIF or Sensitive Information Facility at University of Copenhagen (KU/UCPH) is meant for storing and sharing sensitive data. In particular personal data requiring special care under the EU General Data Protection Regulation (GDPR). The intended audience is employees, their collaboration partners and students. SIF delivers secure centralized storage space for various projects involving sensitive data and integrates safe sharing with other internal or external project participants as well as network drive use from anywhere. SIF is delivered by [SCIENCE HPC Center](#) and is formally for the faculty of Science, but we do allow exceptions. A roll-out to all of UCPH is underway to support the general UCPH [Data Management Guidelines](#).

General Research Data

For non-sensitive research data we recommend instead using the sister facility [ERDA](#), which provides more flexibility and features.

Getting Started

We've gathered information about SIF in a few short intro guides and a more thorough user guide

- [Sign up Intro](#) for UCPH and external users
- [Project Intro](#) about creating, sharing and accessing projects
- [User Guide](#) with general usage

We strongly recommend reading and following at least the two intro guides if you want to use SIF. For any legal questions or permits please contact your [Faculty Secretariat](#).



Creating a project

This is how I do it:

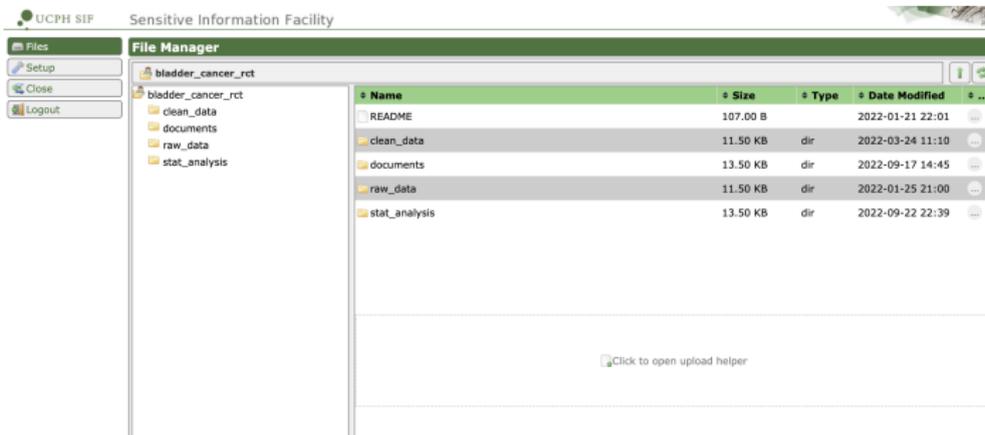
- A general data processing agreement exist between UCPH and Captial Region
- UCPH receives instructions concerning each individual project from the Capital Region
- Instructions are archived in UCPH's system Workzone
- Workzone number is required to create the project in SIF
- Project owner invites collaborators to participate in project on SIF

This is how I feel about it:

- There is a lot of information about Research Data Management and GDPR
- Am I really following the rules?
- People at UCPH help as best they can



Displaying the content of a project



The screenshot shows the Sensitive Information Facility (SIF) web interface. The top navigation bar includes 'UCPH SIF' and 'Sensitive Information Facility'. A sidebar on the left contains 'Files', 'Setup', 'Close', and 'Logout'. The main area is titled 'File Manager' and displays a tree view of a project named 'bladder_cancer_rct'. The tree view shows a folder named 'bladder_cancer_rct' containing subfolders: 'clean_data', 'documents', 'raw_data', and 'stat_analysis'. The main pane shows a table of files and folders with columns for Name, Size, Type, and Date Modified.

Name	Size	Type	Date Modified
README	107.00 B		2022-01-21 22:01
clean_data	11.50 KB	dir	2022-03-24 11:10
documents	13.50 KB	dir	2022-09-17 14:45
raw_data	11.50 KB	dir	2022-01-25 21:00
stat_analysis	13.50 KB	dir	2022-09-22 22:39

Click to open upload helper

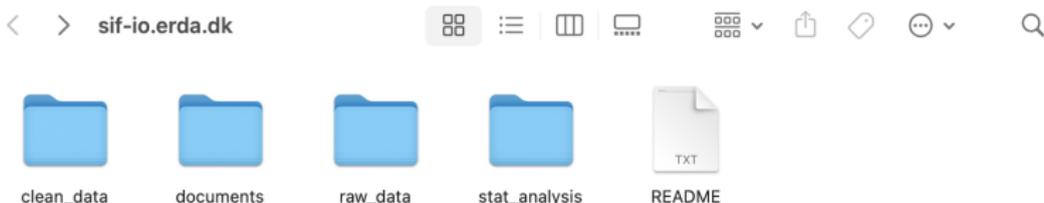
I find the web interface useful for

- project management
- uploading documents and (smaller) files with meta data related to the project

Difficult to avoid that files are downloaded and stored locally!



Working efficiently with data on SIF



- I map SIF project folders to my local PC
- I do data analyses on my local PC as if they were local folders
- My student assistants claim that this is very slow
- We have used WebDAVS as Mac OSX supports this
- Support team behind SIF suggests that I use SSHFS
- I often work on small datasets and it is unclear to me how my approach scales to larger data



Personal view on research collaboration

Efficient research collaboration is characterized by

- transparency
- reproducibility
- inclusion of all collaborators when interpreting results

Principles: **Separate data cleansing from data analyses and use output format integrating code, output and accompanying comments.**



The screenshot shows a web interface for a research project. On the left is a vertical navigation menu with the following items: **Overview**, Indlæsning af data, Efterspurgte tabeller, Completion rate, PRO-CTCAE, Completion of treatment, Hospitalization, Dose reductions, QoL (EHS) over tid, QLQ-C30 over tid, Overlevelsesanalyse, and Figur til artikel. The main content area is titled "Statistisk Analyse af data fra RCT med blærekræftpatienter" by Anders Tolver & Mark Knudsen, dated 22 september, 2022. It includes sections for "Oversigt", "Indlæsning af data", "Kliniske data", "PROs", and "Efterspurgte tabeller". Each section has a "Code" button next to it. The "Efterspurgte tabeller" section includes a sub-section "Tabel 1 (Kliniske data)" with a note: "Tabellen med kliniske data er egentlig udfyldt i Grønt dokument, men data reproducens her for god ordens skyld."



Concluding remarks and questions

- A few "*common use cases*" for working with sensitive data would be helpful
- Having access to a lot of information about storing and processing of sensitive data does not ensure proper use
- Having access to facilities for storing sensitive data does not ensure proper use
- Publishing sensitive data: guides and technical support?
- How to use research based on sensitive data for educational purposes?
- Clarification of concepts like anonymisation and pseudonymisation when metadata may be available elsewhere

