

# DANS DCC Tech Support

## *DANS Long Term Preservation Back-end for Local DCCs*

Jan van Mansum, Core Services Team Leader, DANS

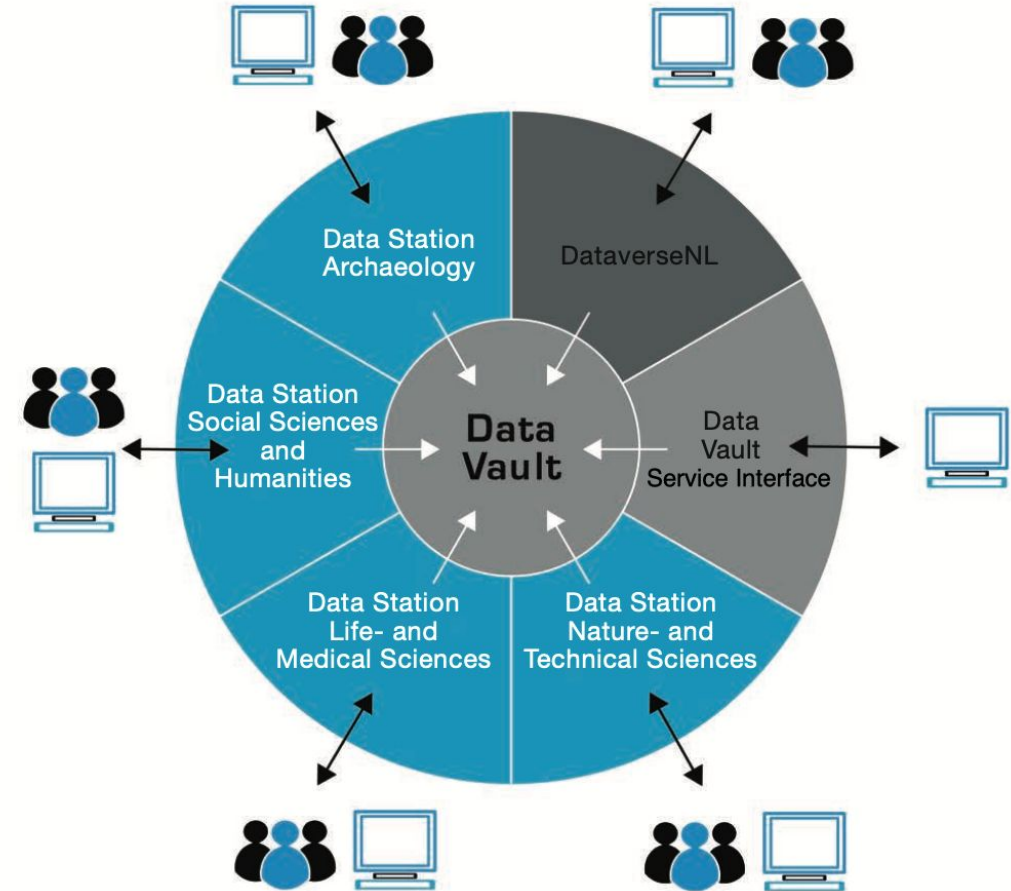
29 September 2022

# Agenda

- Context: EASY → Data Stations
- Project Goals
- How does it work?
- Questions

# Context: from EASY to DANS Data Stations

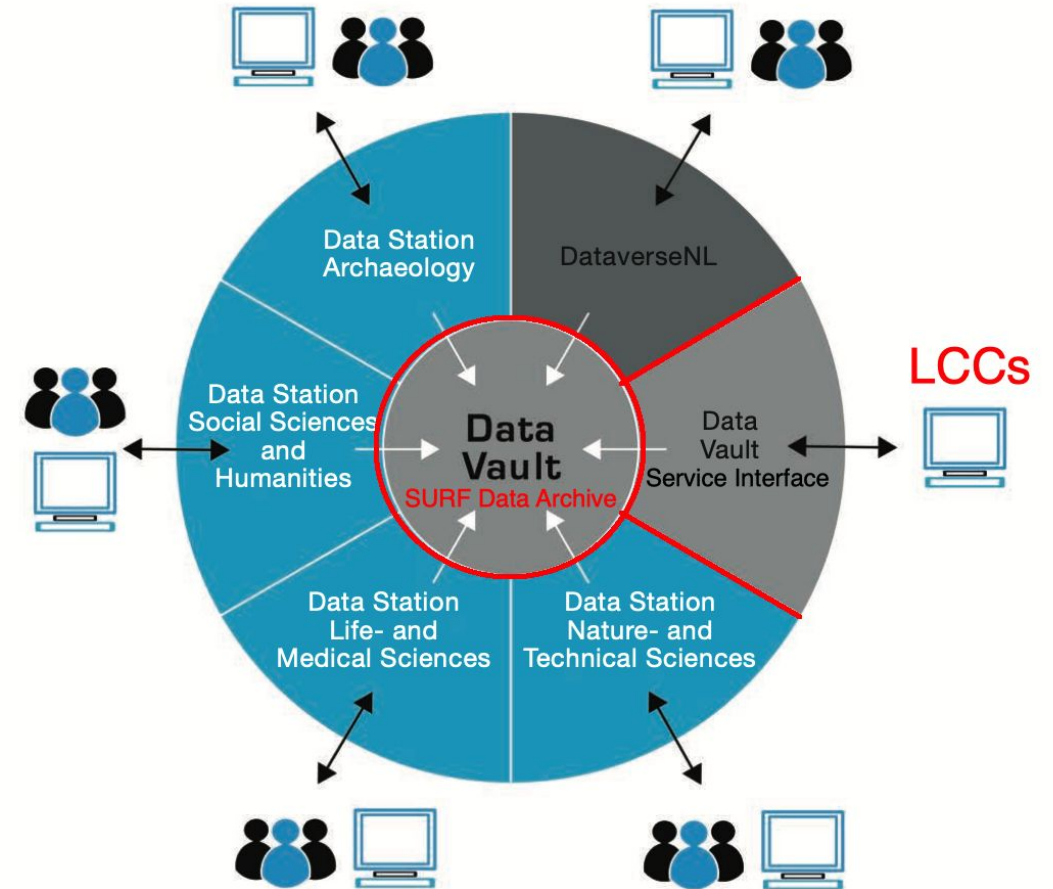
- EASY:
  - since 2007
  - one size fits all (disciplines)
  - in-house developed
  - old technologies
- Data Stations
  - since 2022-...
  - discipline specific
    - + DataverseNL
  - core: Dataverse software
- Data Vault
  - long-term preservation copy
  - SURF Data Archive
  - **has service interface**



# "DCC Tech Support" Project

## Project Goals

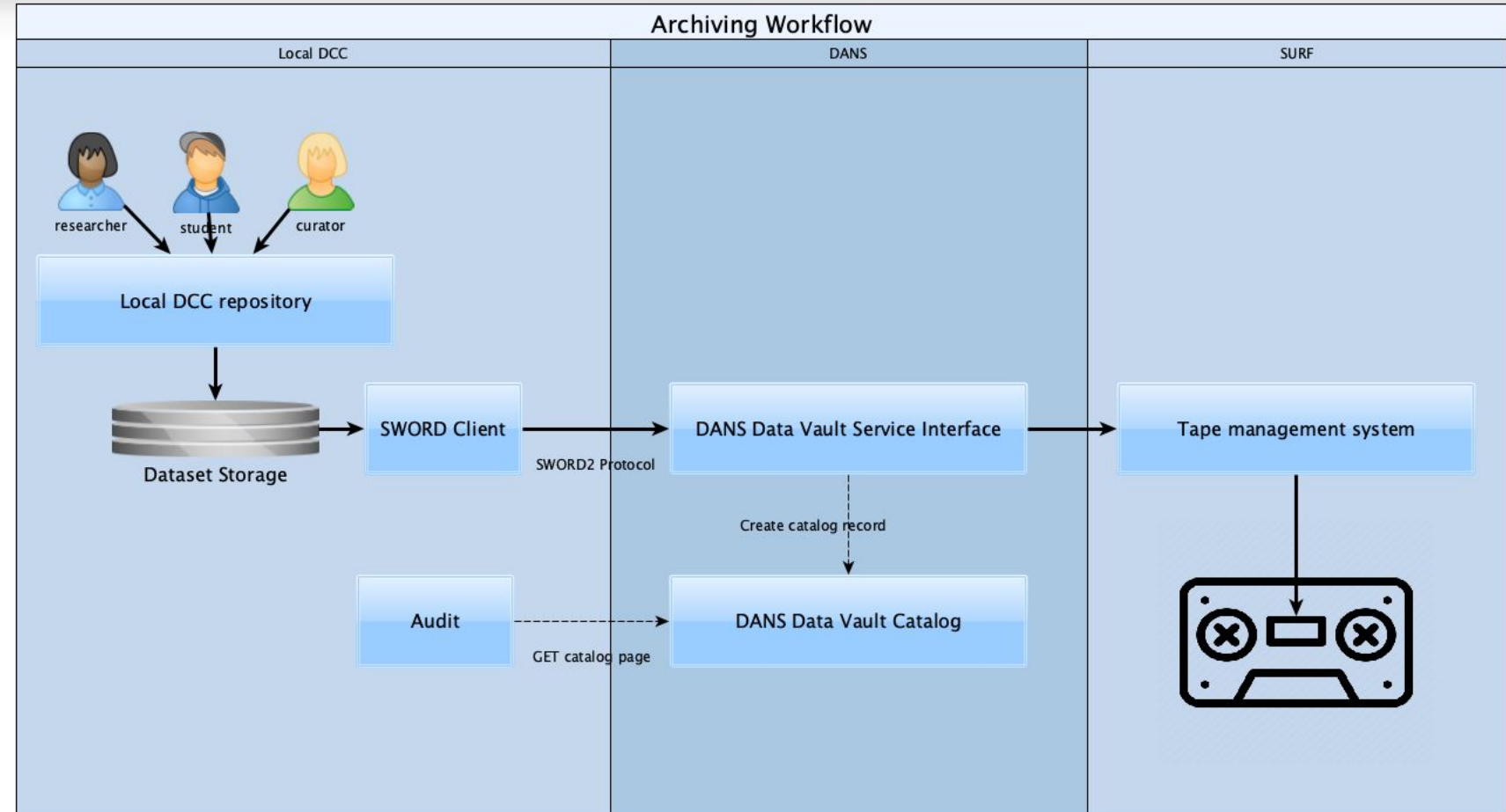
- Give LCCs access to Data Vault
- Via machine-machine interface
- Cooperation with SURF by using SURF tape storage (Data Archive)



# How does it work?

## Features:

- SWORD2 interface
- Dataset can be any\* size
- Archived in software-independent OCFL lay-out
- Recorded in catalog
- Catalog record retrievable through NBN PID



# Next steps

- Take Vault Service Interface to production
- Migrate existing customers from EASY
- Add functionality to service:
  - Dashboard for reporting/statistics

# Questions?



# THUIS: Trust Higher by Using iRODS and SRAM

DCC September, 29 2022







# Goal THUIS

- Set up an architecture and privacy framework to be able to process highly confidential data according to the FAIR principles within the Amsterdam UMC research cloud.
- Building blocks: SURF Research Access Management (SRAM), iRODS, SURFcumulus cloud management en Azure.
- Results should be extendable to SURF ResearchDrive, myDRE (anDREa) and HPC.



# Adapted Goals (1/2)

- Discontinuation of SURFcumulus CMP on January, 1 2023.
- CMP core building block in Amsterdam UMC Research Cloud.
- Stop further development of Amsterdam UMC Research Cloud.
- Alternatives compute facilities needed for highly confidential data.



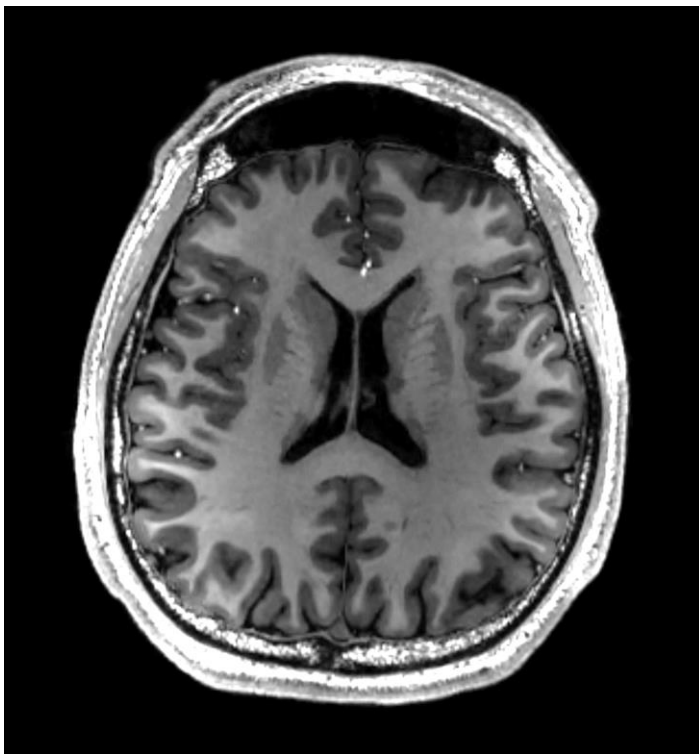
# Adapted Goals (2/2)

Project divided in 3 parts:

- Make the Amsterdam UMC Research Cloud suitable for highly confidential data (stopped) followed by transition of Amsterdam UMC Research Cloud to a new platform suitable for highly confidential data (not part of this project).
- iRODS for highly confidential data.
- iRODS/YoDa for medium sensitive data (added).



# Need for highly confidential data platforms



- Jan C. Mazura et. Al, Facial Recognition Software Success Rates for the Identification of 3D Surface Reconstructed Facial Images: Implications for Patient Privacy and Security (Journal of Digital Imaging, 2011).
- de Sitter, A., Visser, M., Brouwer, I. et al. Facing privacy in neuroimaging: removing facial features degrades performance of image analysis methods. Eur Radiol 30, 1062–1074 (2020). <https://doi.org/10.1007/s00330-019-06459-3>



# Need for highly confidential data platforms

door **Jelle Boonstra**

**ZWOLLE** - Pieter V. mag voorlopig nog als verdachte worden aangemerkt in de dubbele moord op zijn ouders, Marrie en Jurrie Vis, uit Vollenhove. De Raadkamer in Zwolle besloot gisteren om het openbaar ministerie opnieuw drie maanden respijt te gunnen in onderzoek naar de moordzaak uit september 2009 - de officier van justitie maakte volgens de raadkamer van de Rechtbank Oost-Nederland 'voldoende nemelijk' dat onder meer DNA-onderzoek naar sporen van een horloge, de kraan en de pols van de moeder me...





# Amsterdam UMC Research Cloud suitable for highly confidential data

Stopped due to discontinuation of SURF CMP

Results before discontinuation:

- High Level Design (HLD; on request)
- Requirements document research cloud for sensitive data
- Request for Proposal tender

Conclusion: SRAM adds a level of complexity that might exceed the effort.



# iRODS highly confidential (1/2)

Based on use case Human Genetics

- Core-facility Genomics sequenced whole genome (highly confidential).
- Data will be included into iRODS including metadata.
- Used to be able to use metadata on HPC-facilities (e.g. SURF resources)

Dedicated solution for Human Genetics

- iRODS hosting (SURF) for highly confidential data
- Dedicated hardened clients needed for secure connection
- SRAM not available yet within iRODS (yet)
- Data processing agreement (verwerkingsovereenkomst) between SURF and Amsterdam UMC



# iRODS highly confidential (2/2)

## Deliverables:

- Highlevel design iRODS for Human Genetics (HLD; on request)
- Measures needed for iRODS highly confidential (on request)
- Report iRODS highly confidential
- BIV/DPIA





# YoDa/iRODS medium confidential (1/2)

Offer iRODS/YoDa for medium confidential data using an Amsterdam UMC specific metadata schema.

- Separate instance from iRODS for Human Genetics
- Goal getting experience with the YoDa and the Amsterdam UMC specific metadata schema.



# YoDa/iRODS medium confidential (2/2)

## Deliverables:

- Metadata schema Amsterdam UMC
- Evaluation YoDa in combination with metadata schema

## Notes:

- 2 FA wasn't available up to now for YoDa (will be soon)
- All universities are part of YoDa consortium and RDM tech group, but until now no UMC's.
- UMC's should strive for a platform for metadata management during projects