Data Management Plan Seminar, LUX May 7 2019

DMP - why have one, and why should you care?

Anders Sjöström
LUNARC/SNAC

What is SNIC/SNAC?

## SNIC – Swedish National Infrastructure for Computing

SNIC - a national research infrastructure that makes available large scale high performance computing resources, **storage capacity**, and advanced user support, for Swedish researchers.

SNAC – Swedish National Allocations Committee

SNAC - a body within SNIC which independently is responsible for evaluating proposals for access to the SNIC resources and for assigning allocations for SNIC LARGE.

My Background
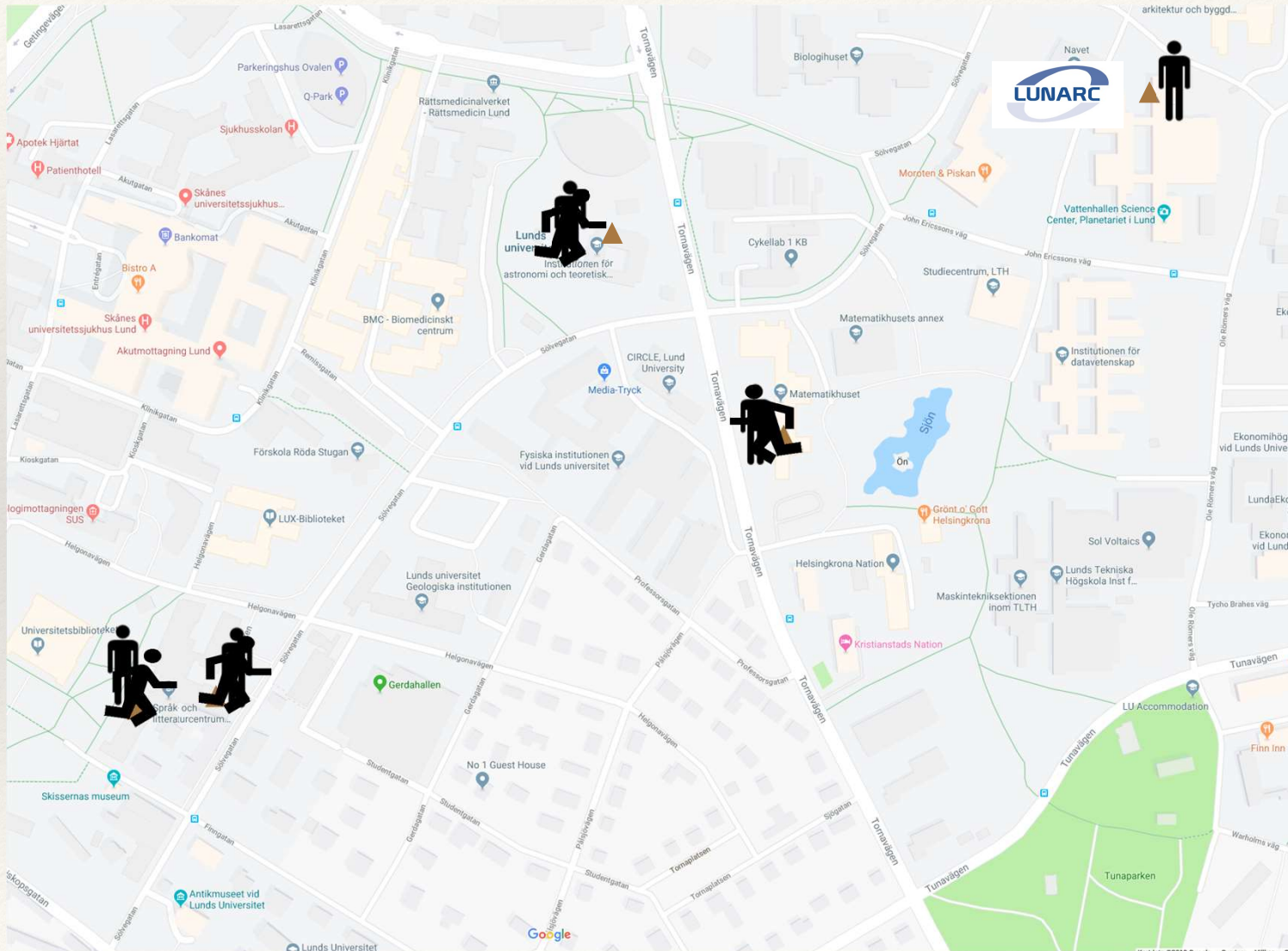
My road to data management enlightenment

Perhaps focus on the data rather than the production of it?

Ask yourself:

- What is my research question?

- What type of data will I need?

- How much data will I need?

- What type of information will I need about the data?

- How do I keep the data up-to-date?

- How long will I need to keep the data?

- How do I move the data?

- Is my data sensitive, and who should have access to it?

- Should the data be published? If so, how?

Wishlist:

- Store data automatically

- Store data with metadata

- Couple a lab journal, either in analogue or digital form to the stored data

- Store the data with backups

- Easy to add and subtract data

- Version control

- Easy access to, and export of data

- Data security and access control

"New" requirements

- "Policies on open access to scientific research results should apply to all research that receives public funds."

  (2012/417/EU)

- "Forskningsresultat, som forskningsdata och vetenskapliga publikationer, som tas fram med offentlig finansiering bör vara öppet tillgängliga så långt det är möjligt."

  (Regeringens proposition 2016/17:50, p.107)

- The Archives Act - *(Arkivlagen SFS 1990:782)*
- GDPR - *(EU) 2016/679*
- Public Access to Information and Secrecy Act - *(offentlighets- och sekretesslagen SFS 2009:400)*
- Research ethics *(SOU 1999:4)*
- Information security
- FAIR & OpenAccess policies

- Requirements from research funding bodies
- Requirements from scientific journals

More data will need to be managed

Data: "information, in particular facts or numbers, collected to be examined and considered as a basis for reasoning, discussion, or calculation." *(Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020. p.4)*

"as open as possible, as closed as necessary". *(ibid. p.8)*

- Data need to be stored
- Metadata need to be stored
- Data will need to be validated
- Data will need to be traceable
    - Creation/Change
    - Security
    - Access
- Data will need to be purged
- Data will need to be open and secure

Where should the data be managed?

Where should the data? / Who owns the data?

In most cases the municipality, more precise: / Locally in the department?
the administrating organisation (medelsförvaltaren) / Centrally within the University?
- Nationally?
- EU?

So what is the task for SNIC/SNAC?

SNIC has been tasked by VR to report the usage of data storage at the SNIC computing facilities

The reporting to VR shall include (but is not limited to) the utilisation of the resource, the number of users, the Field of Research, and gender

*"In order to meet all the requirements and needs from different levels of users to the financiers we need a common policy for the different types of storage … storage that is available **within** a SNIC center."*

To facilitate reporting and monitoring, SNIC has decided to differentiate between projects for Computation and projects for Storage
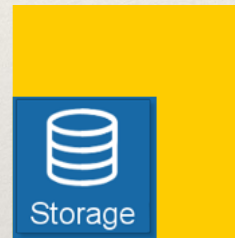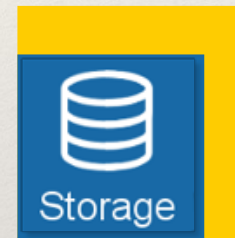
Project storage at SNIC/SNAC?

SNIC has decided to implement project storage in
three volume tiers:
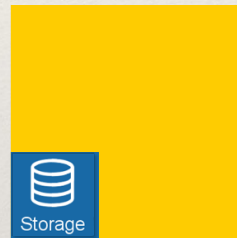
SMALL	MEDIUM	LARGE

For each tier there will be a requirement for a DMP
Note that for SMALL and MEDIUM the DMP can be abridged

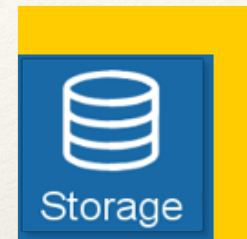# So what is expected in DMPs for project storage?

SMALL

- Short description of the type of data to be used/created
- Short description of any tools to be used to handle/create data
- Short description on how data is to be handled when it may no longer be stored on the resource

MEDIUM

- Description of the type of data to be used/created
- Description of flows of data, including tools and processing of data
- Description on how data is to be handled when it may no longer be stored on the resource

LARGE

- Detailed description of the data to be stored. Including estimated size (GB), estimated number of files, timespan (months), and format.
- Detailed description of the type of data to be used/created
- Detailed description of flows of data, including tools and processing of data
- Detailed description on how data is to be handled when it may no longer be stored on the resource
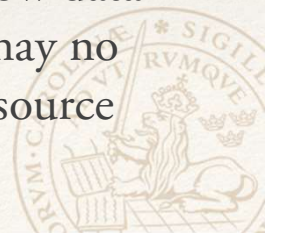
So what is the view of SNIC/SNAC?

- DMPs are beneficial for both the users of SNIC resources and for SNIC

- SNIC will only require a subset of a full DMP in the application process for project storage

- SNIC assumes that any PI will have a full DMP which can be referenced to extract the information needed in the application of project storage

- As project storage is temporary in nature it is required from the PI to provide a plan on how the produced/stored data is to be managed once the project storage is no longer available

My personal remarks/conclusions

- DMPs are good for you

- A properly used DMP will be a support and guide in your research project

- Do NOT "write and forget", update the DMP whenever you make changes in the way you gather, process, create, or purge data

- If you make changes to your DMP, make a note of when and why you changed it and archive the previous version

The DMP is your friend, trust the DMP *Paranoia 1986*