



Swedish National Data Service

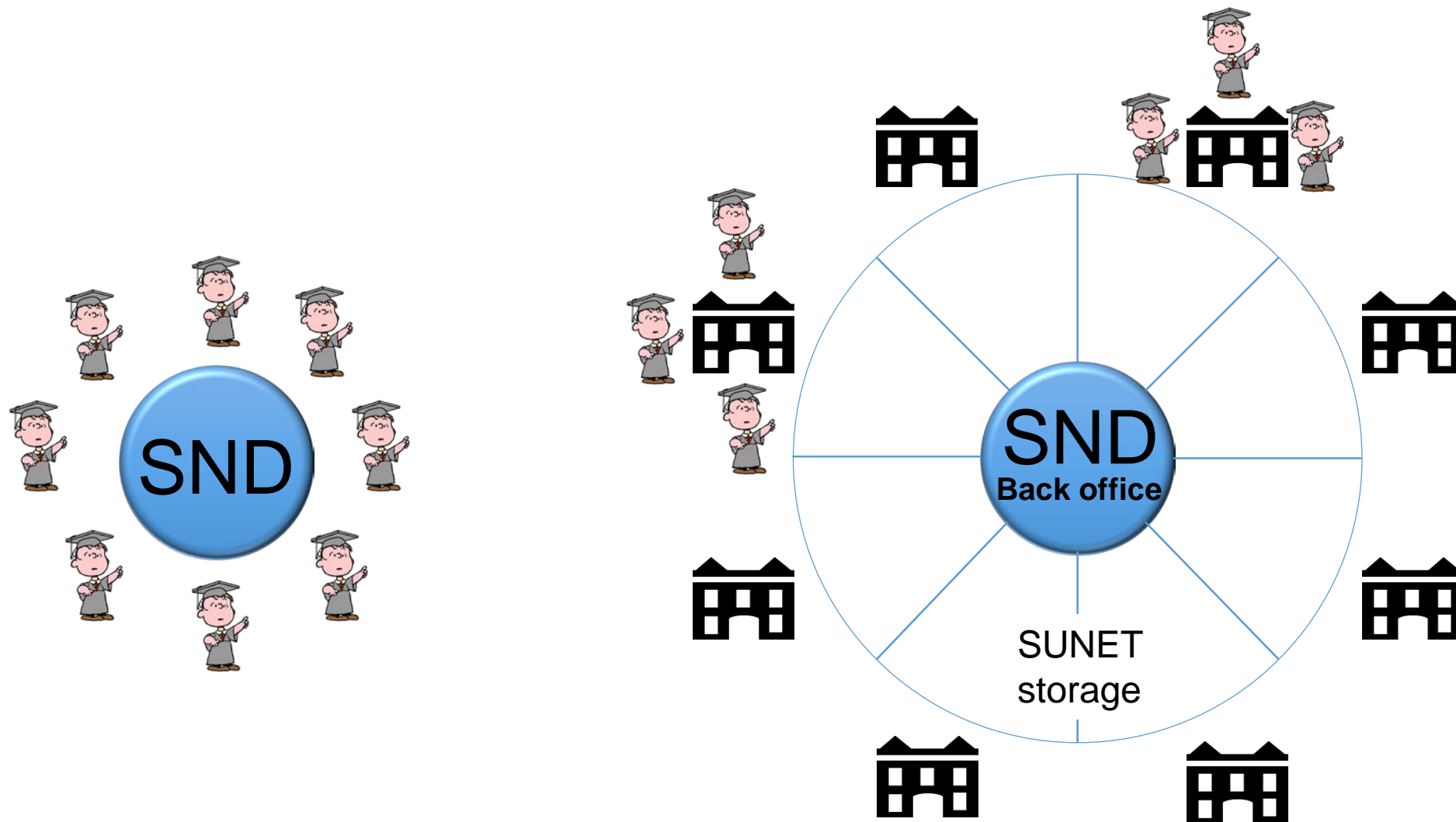
An infrastructure for research data

Max Petzold, director





Remember SND 1.0 vs SND 2.0





Freely Accessible Data

1. The researcher and DAU enter metadata into "My SND" and curate the data.



Metadata are uploaded to the SND portal.

Data are uploaded to the university's partition in the national storage. Flagged as open and made directly accessible.



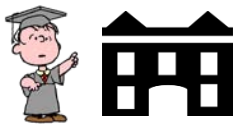
2. Another researcher finds the data in the portal and downloads directly from storage via "My SND".





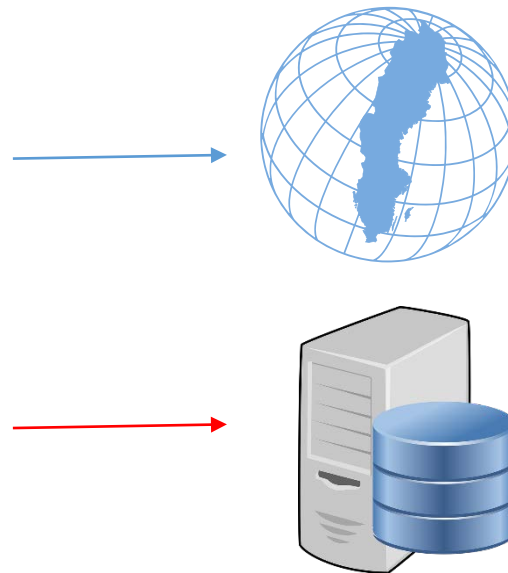
Restricted Data

1. The researcher & DAU enter metadata into "My SND" and curate the data.

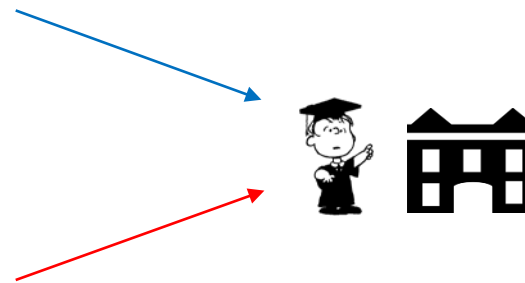


Metadata are uploaded to the SND portal.

Data are uploaded to the university's partition in the national storage. Flagged as restricted and not directly accessible.



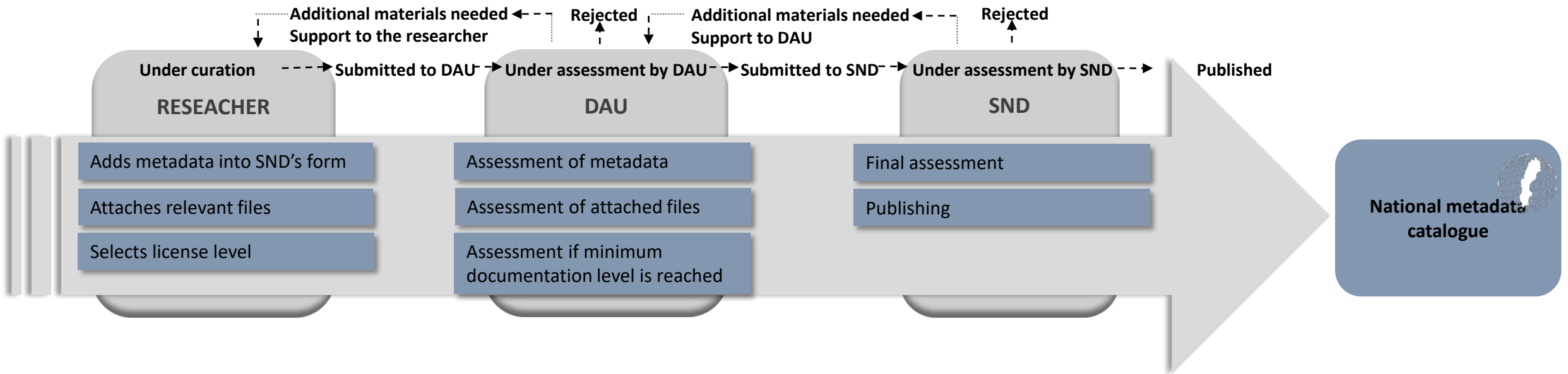
2. Another researcher finds the data in the portal and an access request is sent to the DAU. After the request is cleared by the university (license and/or legally) a download link is sent to the researcher in "My SND".





Workflow: Researcher – DAU – SND

Description below is for the first period and might be revised over time



Research data stay safely with the owner university in the national storage

Långtidslagring av forskningsdata

Gemensam lösning för långtidsförvaring
2018-10-08

FAIR Data principles

TO BE FINDABLE

- F1. (meta)data are assigned a globally unique and eternally persistent identifier.
- F2. data are described with rich metadata.
- F3. (meta)data are registered or indexed in a searchable resource.
- F4. metadata specify the data identifier.

TO BE ACCESSIBLE

- A1 (meta)data are retrievable by their identifier using a standardized communications protocol.
 - A1.1 the protocol is open, free, and universally implementable.
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
- A2 metadata are accessible, even when the data are no longer available.

TO BE INTEROPERABLE

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles.
- I3. (meta)data include qualified references to other (meta)data.

TO BE RE-USABLE

- R1. meta(data) have a plurality of accurate and relevant attributes.
 - R1.1. (meta)data are released with a clear and accessible data usage license.
 - R1.2. (meta)data are associated with their provenance.
 - R1.3. (meta)data meet domain-relevant community standards.

Bygga yta för långtidslagring

Egen

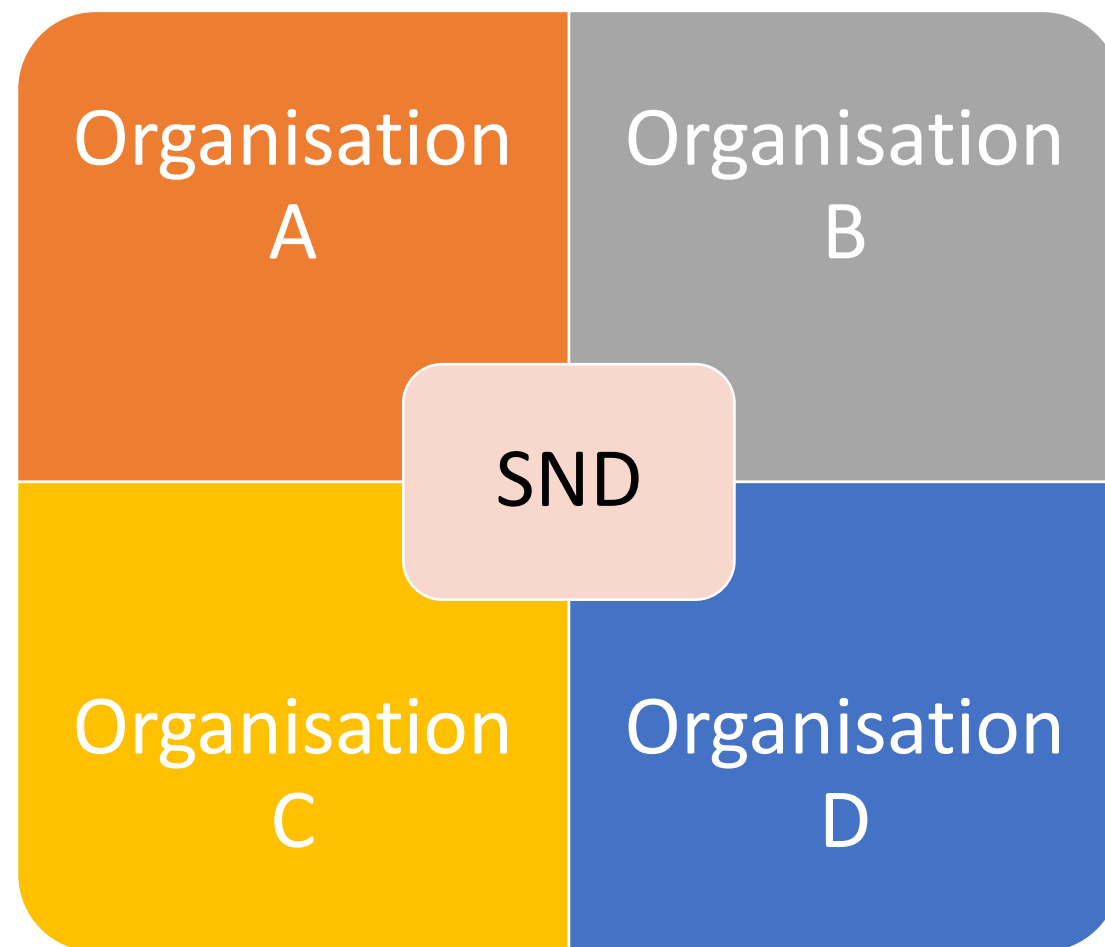
- Kan komma igång snabbare
- Enklare att bygga upp
- Egen upphandling av hårdvara
- Geografisk replikering möjligt men kräver avtal med leverantörer

Gemensam

- Kostnadseffektivt
- Geografisk replikering
- Driftsäkerhet centraliserat
- Mindre organisationer får samma möjligheter som de större

Långtidslagring av data på SUNET

- Projektstart under hösten 2018
- SND:s roll kommer vara att hantera identifiering och metadata
- Varje organisation kommer ha sin yta
- Kort statusuppdatering från SUNET för uppstarten:
 - Oktober: leverans och bygge av första noden i Stockholm
 - November: leverans och acceptanstester
 - December: första kunder live i noden i Stockholm



Vad, när hur?

- Lösning med SUNET har flera fördelar om flera större universitet samarbetar
- Än så länge är det inte helt fastställt hur det ska fungera
 - Input från flera lärosäten
 - Kommer byggas på fler funktioner längre fram
- SND deltar i referensgruppen för den distribuerade lagringen
 - Möte i slutet av oktober
- SND kör igång tester mot lagringsytan på SNIC under hösten

Långtidslagring = Arkiv?

Kan en gemensam långtidslagring fylla kraven för arkivering?

Krav:

- Säker drift
- Rutiner för backup
- Metadata
 - Identifiering
 - Preservering
- Rutiner för kontroller av data

Om långtidslagringen kan fylla kraven för arkivlagring underlättas mycket för universiteten

Långtidslagring : tillgänglighet

- Tillgång till öppna data
 - Ska vara direkt tillgängligt utan någon form av inloggning för alla
- Data som kräver tillstånd
 - Tillgänglig för användare inom t.ex. EDUGAIN som söker tillstånd från forskaren/lärosätet
- Känsliga data
 - Tillgång först efter t.ex. etikprövning, skrivet avtal etc.