Building a Dutch National Research Infrastructure

IRODS UGM 2017

Frank Heere
15-06-2017
SURF: who are we?

Not-for-profit cooperative for ICT in Dutch education and research

✓ Knowledge sharing
✓ Shared digital infrastructures and services
✓ ICT marketplace for acquiring software, cloud services and digital content
Members include

- 14 Research universities
- 8 Academic Medical Centers
- 34 Universities of applied sciences
- 6 Research institutions
- Many other research and education institutes
SURF: operating companies

**SURFmarket**
- negotiates with ICT providers on behalf of institutes connected to SURF

**SURFnet**
- supports, develops and operates advanced, reliable and interconnected ICT infrastructure networks

**SURFsara**
- Netherlands' national supercomputing centre, supplies high-performance computing services, data storage, network research and visualizations
- Hub to EU infrastructures
Expertise | Consultancy | Training | Visualisation | Optimisation

High Performance Computing
- Supercomputing
- Cluster computing
- HPC Cloud

Data processing
- Data analytics
- HPC Cloud
- Grid services
- Visualisation

Data services
- Mass Online Storage
- PID service
- Data management
- Preservation

Innovation & Collaboration
Research Data Management

RDM is in top 3 priorities of Dutch universities and Academic Medical Centers

- Drivers: reproducible research, and reuse of data
- A Data Management Plan is mandatory in funding proposals in the Netherlands
  - Including obligation of 10 year research data retention period
- EU requires FAIR Open Data
- EU General Data Protection Regulation (privacy sensitive data)
  - In force by May 2018
  - Traceability of data
Typical situation

- Research Data stored on many different storage systems + USB sticks/harddisks
Life Cycle Research Data Management

Ideal situation

Research Question

Data Search / Reuse

Repositories

Data Management Plan

HPC infrastructure

Collection

Data Storage

Description

Analysis

Publication

Repositories

Archive

HPC infrastructure
Advantages for SURF members

✔ Lower / shared development and service costs
✔ Share knowledge, increase community
✔ Develop generic tooling & functionalities that can be used by all institutes
✔ Easier data sharing between universities
✔ Enlarge storage infrastructures, reduce cost per TB
  - Only invest once
✔ Store data close to HPC infrastructure of SURFsara
  - Faster data analysis
National approach on Research Data Infrastructure

Strategy
✓ National Coordination Point Research Data Management
  ▪ Facilitates a national strategy for research data management in the Netherlands

Infrastructure Community
✓ U2CONNECT
  ▪ Netherlands based community advancing and promoting interoperability and connectivity between research data infrastructure services

Infrastructure
✓ Dutch National Research Data Infrastructure
  ▪ To set-up a sustainable infrastructure for the coming 5 to 10 years
National approach on Research Data Infrastructure

Strategy
✓ National Coordination Point Research Data Management
   ▪ Facilitates a national strategy for research data management in the Netherlands

Infrastructure Community
✓ U2CONNECT
   ▪ Netherlands based community advancing and promoting interoperability and connectivity between research data infrastructure services

Infrastructure
✓ Dutch National Research Data Infrastructure
   ▪ To set-up a sustainable infrastructure for the coming 5 to 10 years
Why iRODS?

- RDM is a must for universities, they have to act now, and iRODS seems most viable solution
  - 3 Dutch universities have already developed a solutions based on iRODS
  - Several other universities, medical hospitals and large research projects explore iRODS technology
- Automated workflow implementation for researchers
- iRODS covers GDPR aspects
  - Traceability
  - Transparency
  - Privacy by design
- Obligation of 10 year data retention period
Why NOT iRODS?

- Empty framework
  - Implementation requires huge effort
- No presentation, basic UI layer
  - Customized solution needs to be developed and maintained
- No broad community with iRODS knowledge
  - Risk on continuity
  - Active community building required
- Risk on global support and development iRODS
  - Will it fit the needs of Dutch institutes?
National iRODS expertise center

✓ SURF intends to establish a National iRODS expertise center
  ▪ Sustained availability of knowledge
  ▪ Offer Consultancy, Development & Support
  ▪ Develop a library with standard(ized) tooling, functionalities
  ▪ Develop a presentation layer/UI template
  ▪ Support future iRODS developments
  ▪ Provide iRODS as a Service
    ▪ Middleware layer (both iCAT and resource servers)
    ▪ Flexible storage solutions

✓ Use iRODS to seamlessly connect to the Dutch National & EU Research Infrastructures
iRODS as a Service: Proof of Concept

Proof of Concept architecture

✓ Data replication
  ▪ Copy of data at SURFsara
  ▪ Permanent copy: disaster recovery
  ▪ Temporary copy: Bring data closer to compute facilities

✓ Storage scale-out
  ▪ Data is only located at SURFsara
Proof of Concept architecture

✓ Data replication
  - Copy of data at SURFsara
  - Permanent copy: disaster recovery
  - Temporary copy: Bring data closer to compute facilities

✓ Storage scale-out
  - Data is only located at SURFsara
Conclusions & next steps

- Establish National iRODS expertise center
  - Provide expertise and infrastructure
  - Increase expertise resource pool
    - SURF already hired additional 2 fte for iRODS
  - Infrastructure technical feasibility study is done
- Continue with pilots and PoC’s to support further development
  - your invited...!
- Building service plan
- Establish sustained funding model
Thank you!

Universiteit Utrecht

More info @ frank.heere at surfsara dot nl / +31 6 2673 2843