

RESEARCH DATA MANAGEMENT INNOVATION & SERVICE DEVELOPMENT @ SURF



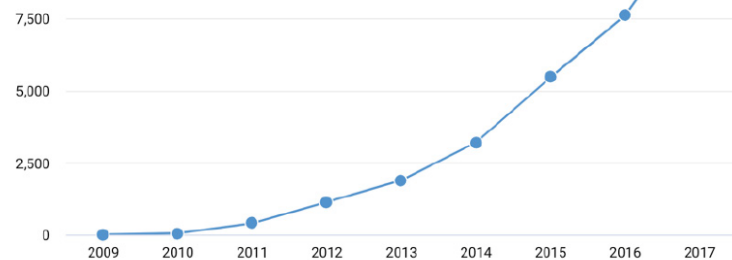
SURF

Wat is Research Data Management?

“Research data management is an explicit process covering the creation and stewardship of research materials to enable their use for as long as they retain value.”

Digital Curation Center

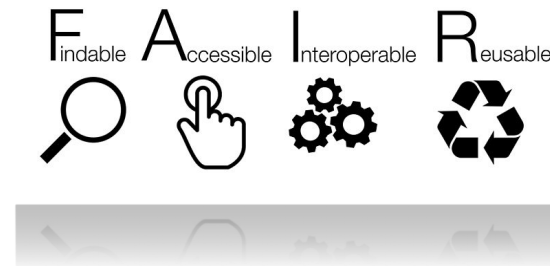
The hockey stick graph indicates the exponential growth of datasets that are being made available.



Source: *The State of Open Data 2018*, Digital Science Report

“Good data management is not a goal in itself, but rather is the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse by the community after the data publication process”

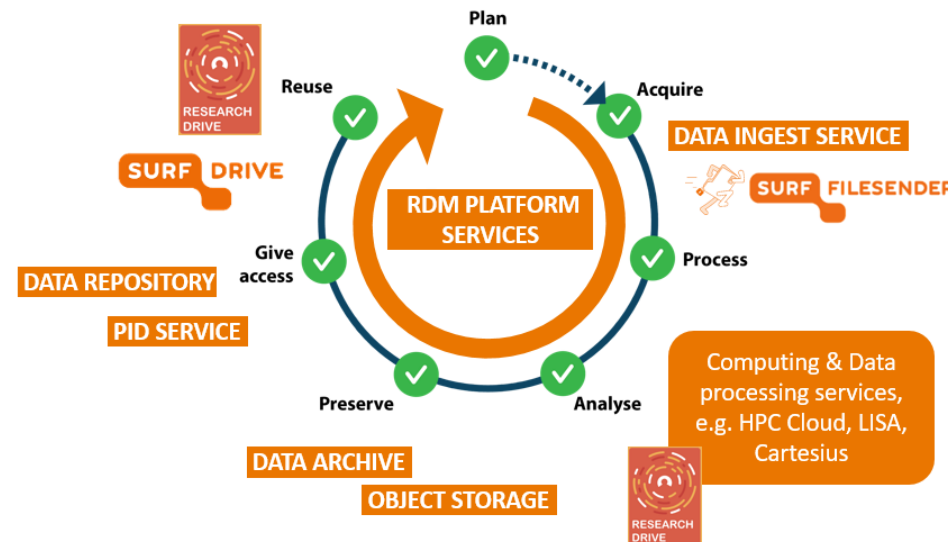
The FAIR Guiding Principles for scientific data management and stewardship



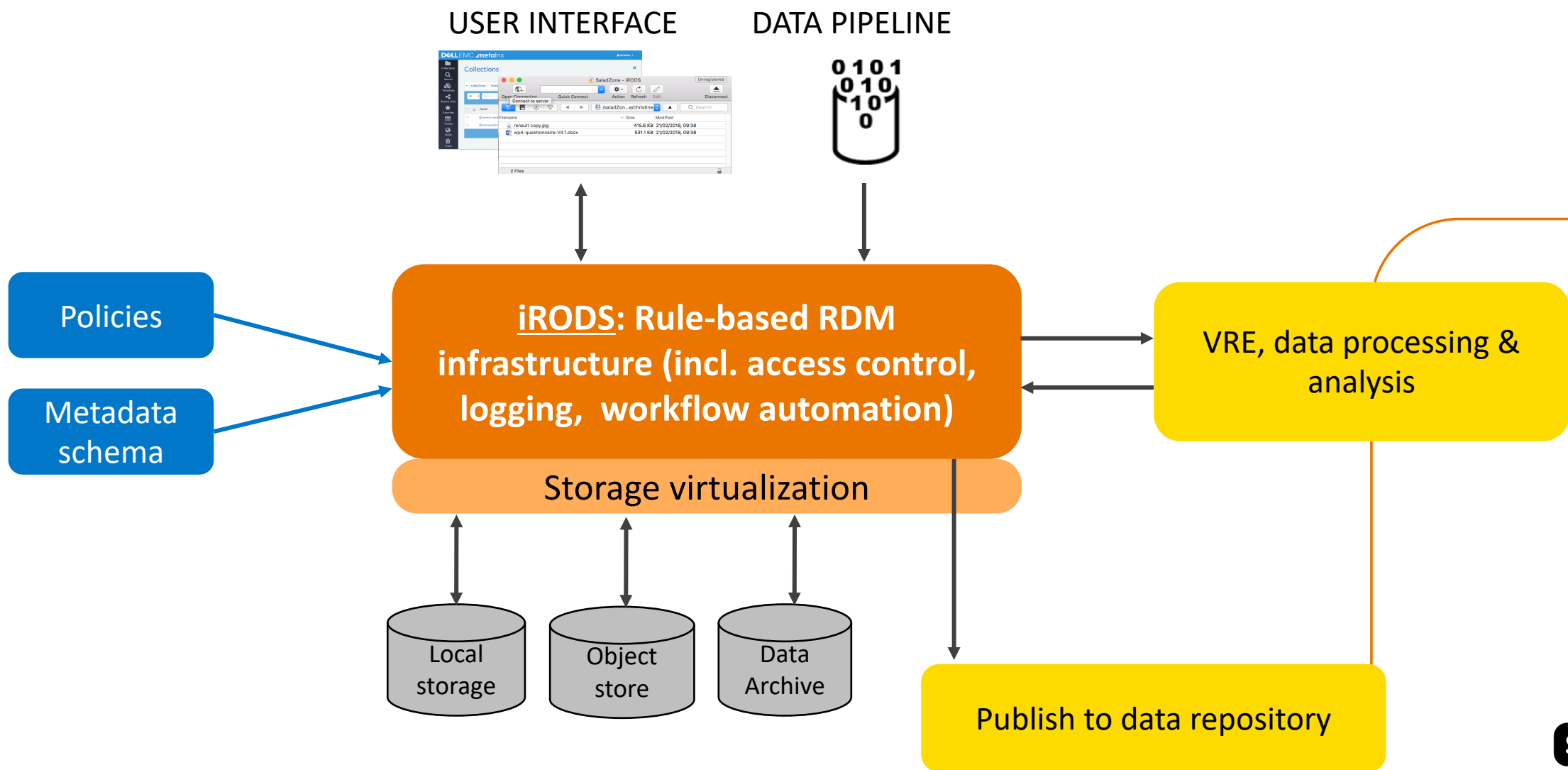
Hoe kan SURF haar leden helpen met het verwezenlijken van hun ambities op (FAIR) RDM gebied?

- **Integrale** aanpak: volledige data lifecycle, diverse actoren (onderzoeker; IT support; bibliothecaris; data steward; etc.)
- Gezamenlijke oplossingen voor gezamenlijke problemen (**synergie**)
- **Integratie** met bestaande SURF diensten
- **Maatwerk** nodig vanwege integratie in lokale context
- **Diversiteit** in kennisniveau en ontwikkel-ambitie

Herbruikbare modules
binnen een algemeen RDM
framework



Dienstontwikkeling: RDM modules met iRODS als spin-in-het-web



Waarom iRODS?

- **iRODS - Integrated Rule-Oriented Data System - is data management middleware:**
 - Data virtualisatie (ontkoppelen fysieke en logische opslaglocaties)
 - Rule-based workflow automatisering
 - User management en access control
 - Metadata management (attribute-value-unit triples)

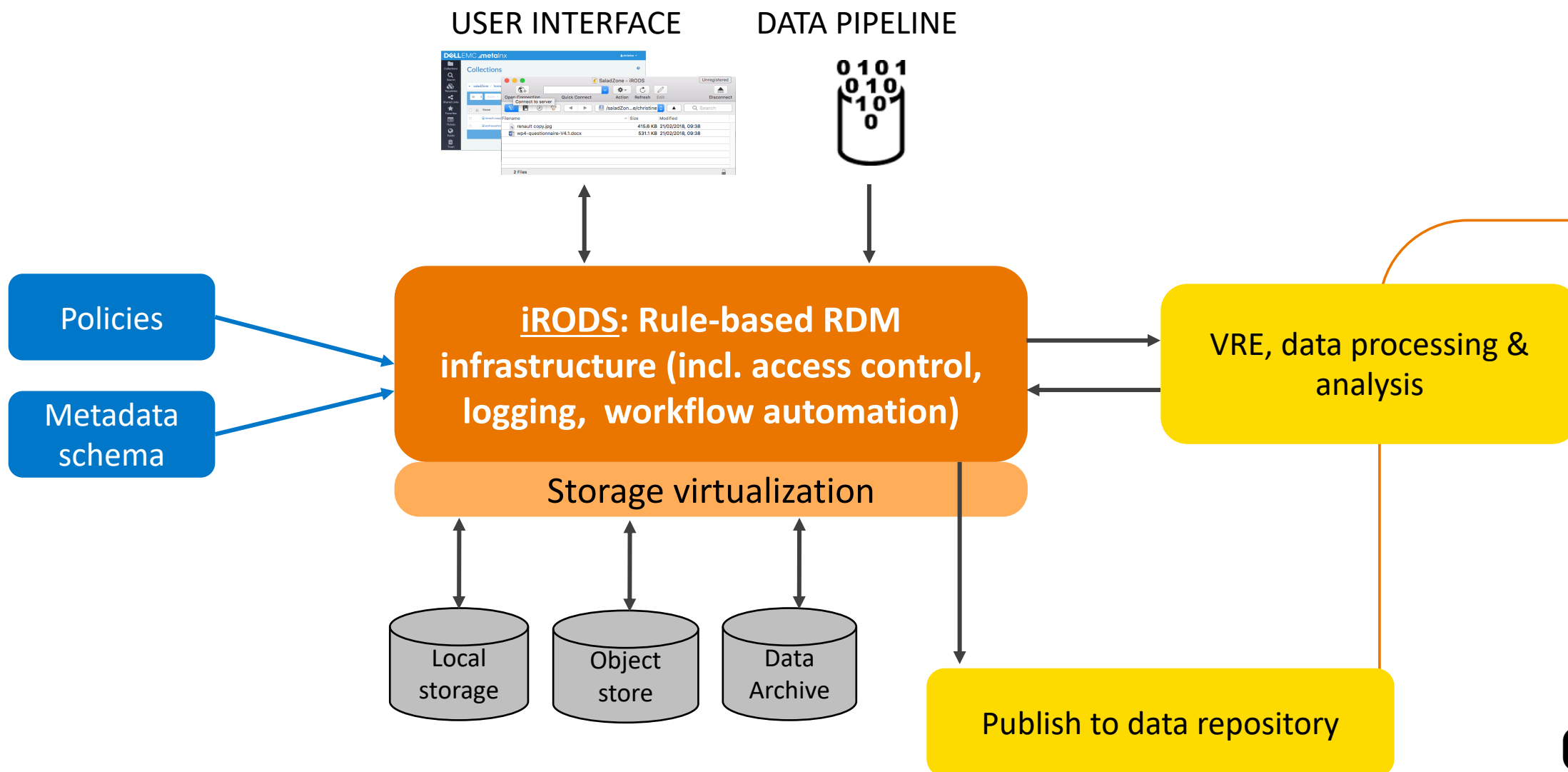
iRODS

Kracht	Zwakte
Performance & scalability	Niet turn-key
Flexibel, integratie door APIs	Niet bedoeld voor gestructureerde data in databases
Open-source; onderhouden door consortium van leden	Geen standaard GUI

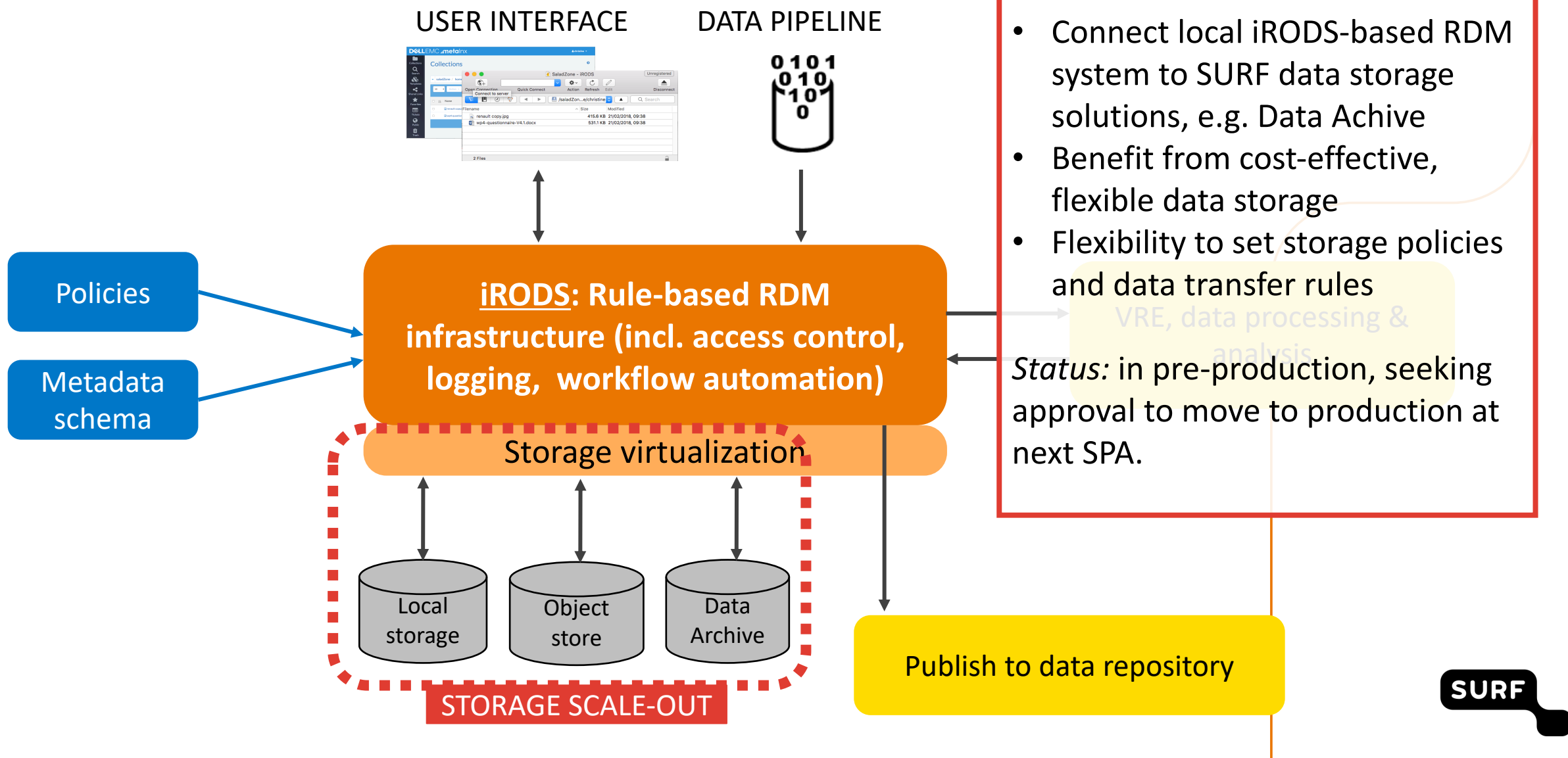
Voorbeelden van iRODS in gebruik:

- NIH NIEHS Data Commons
- EMIF (European Medical Information Framework)
- Utrecht University
- Maastricht MUMC+ DataHub
- RU Groningen
- ... *(en vele anderen)*

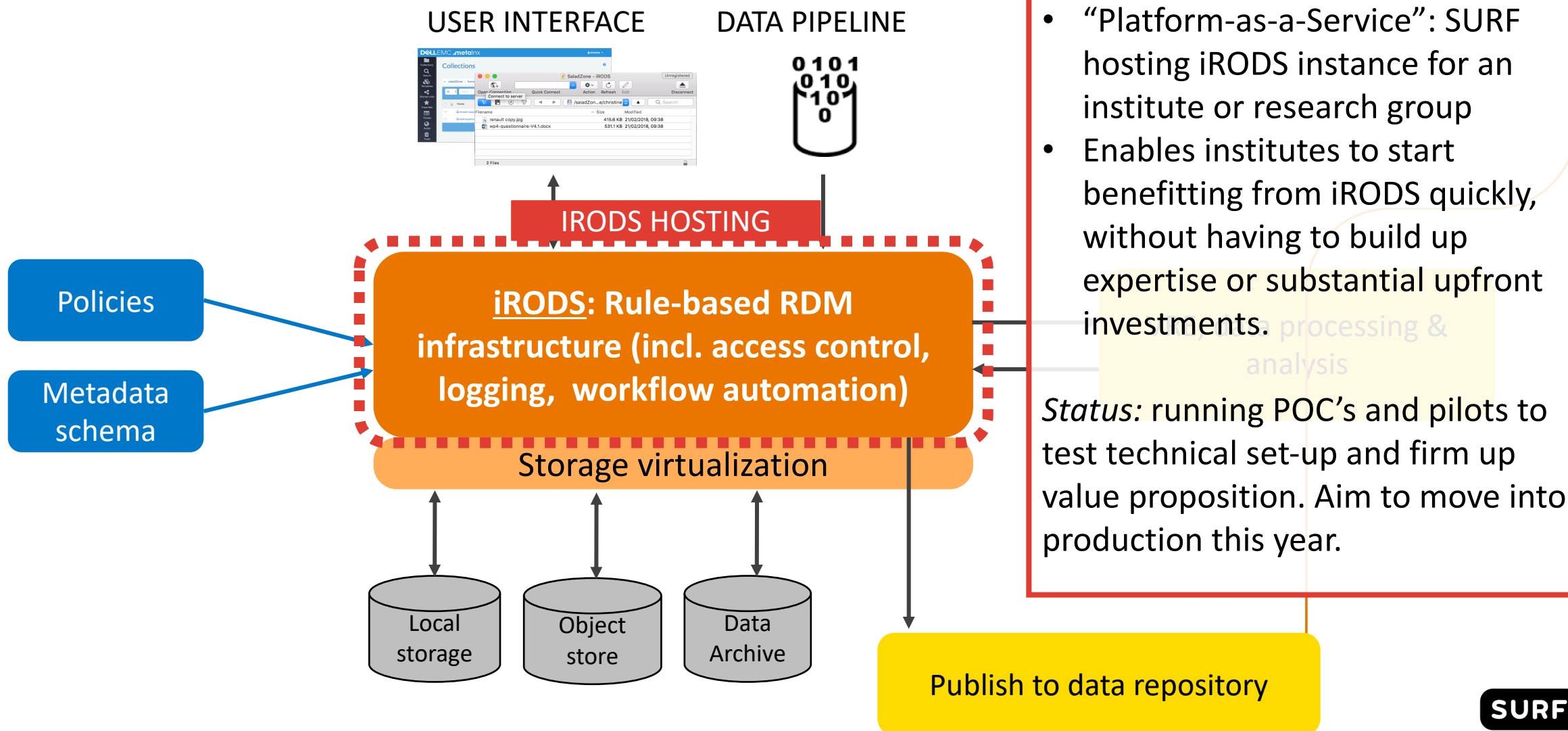
Dienstontwikkeling: RDM modules met iRODS als spin-in-het-web



Module: RDM storage scale-out service



Module: iRODS hosting

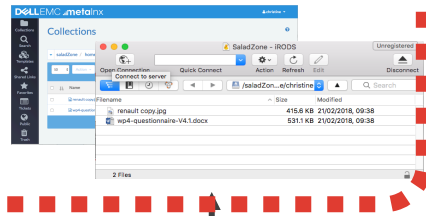


Volgende modules (nog te prioriteren)

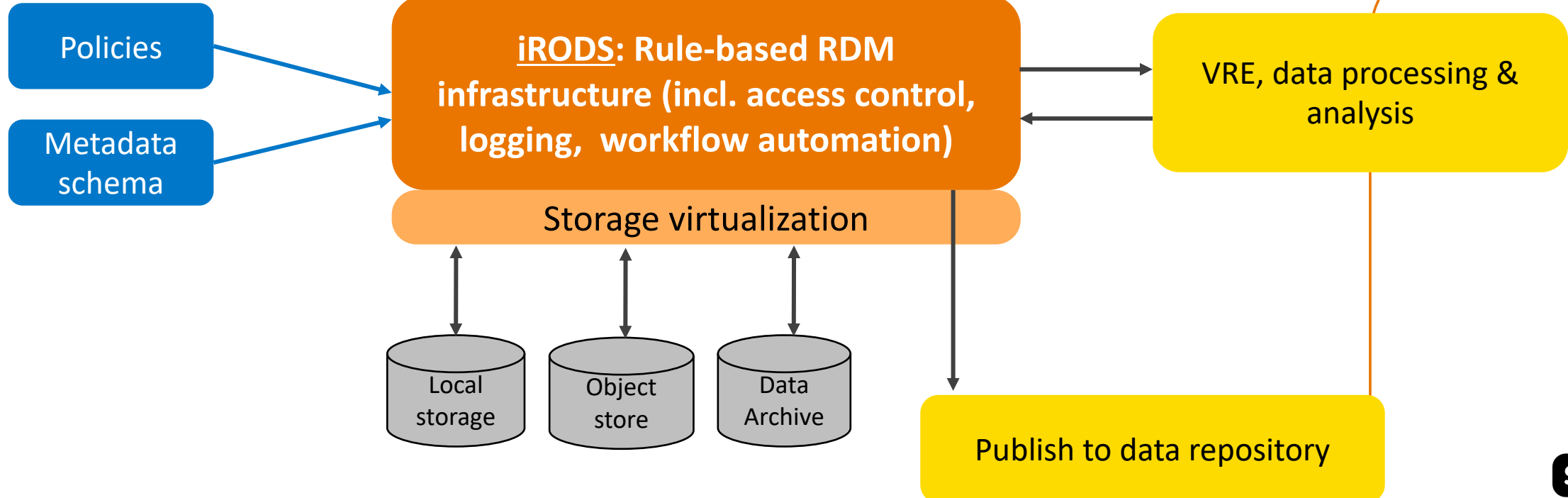
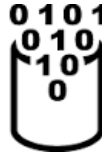
USER INTERFACE:

- YODA (UU)
- KOPPELING MET RESEARCHDRIVE

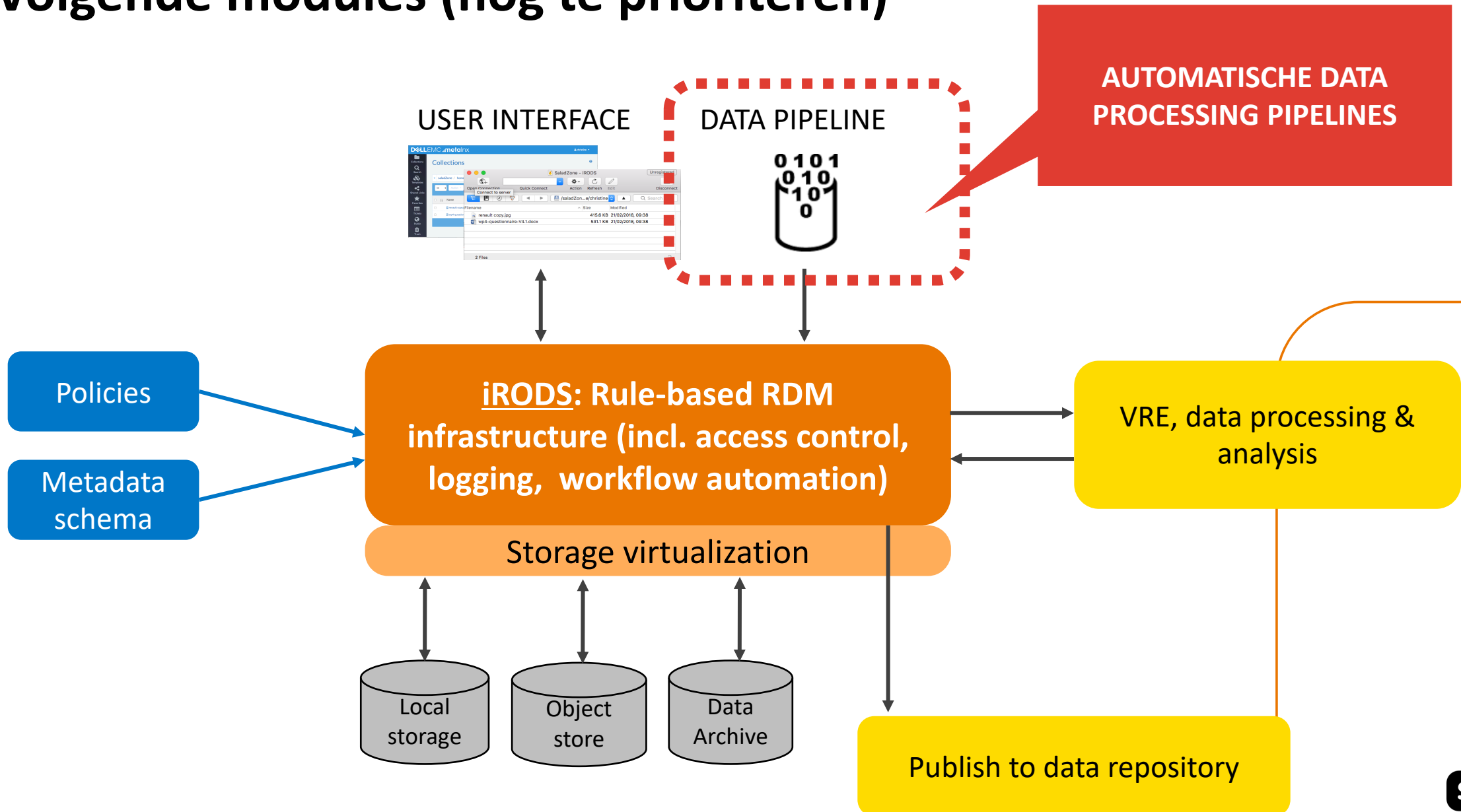
USER INTERFACE



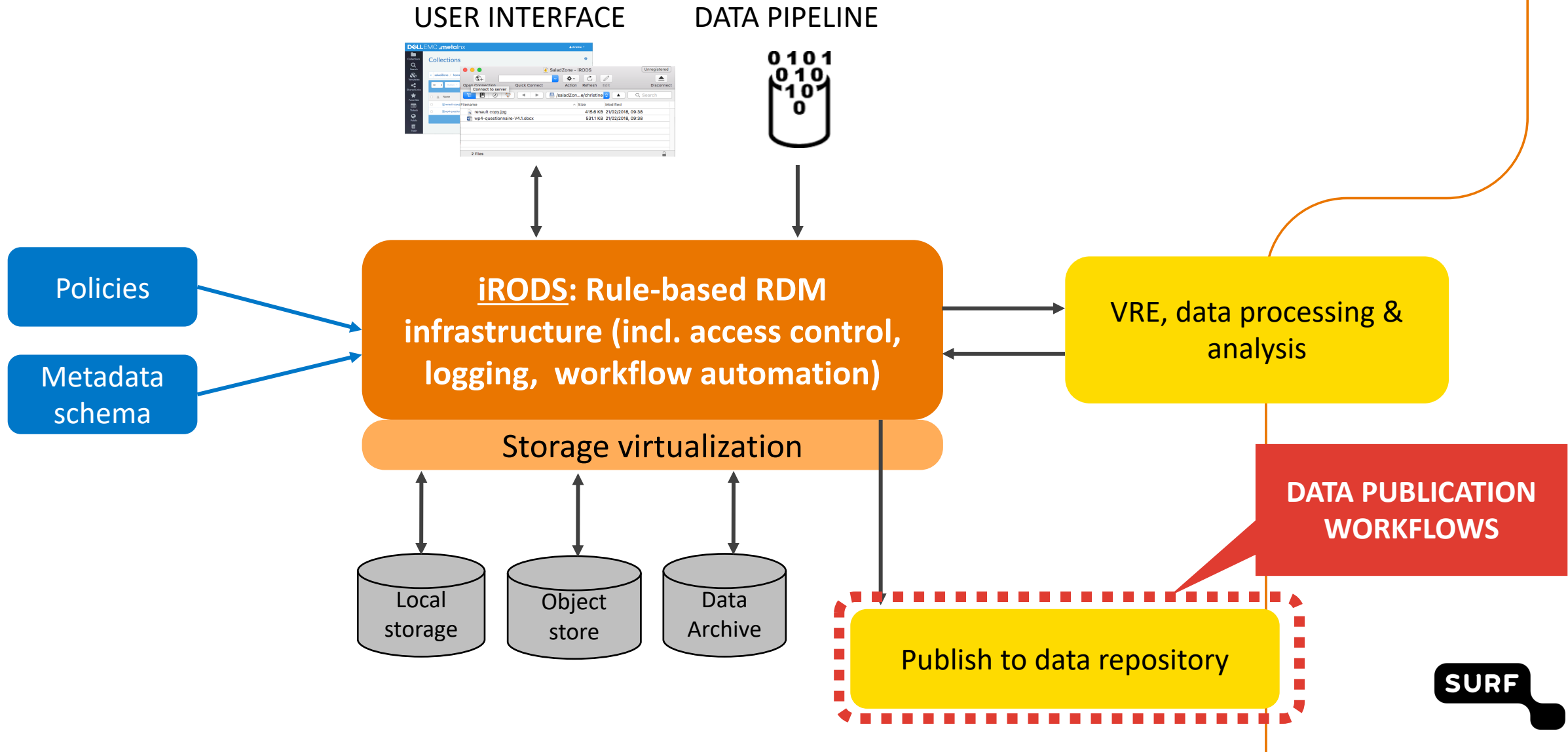
DATA PIPELINE



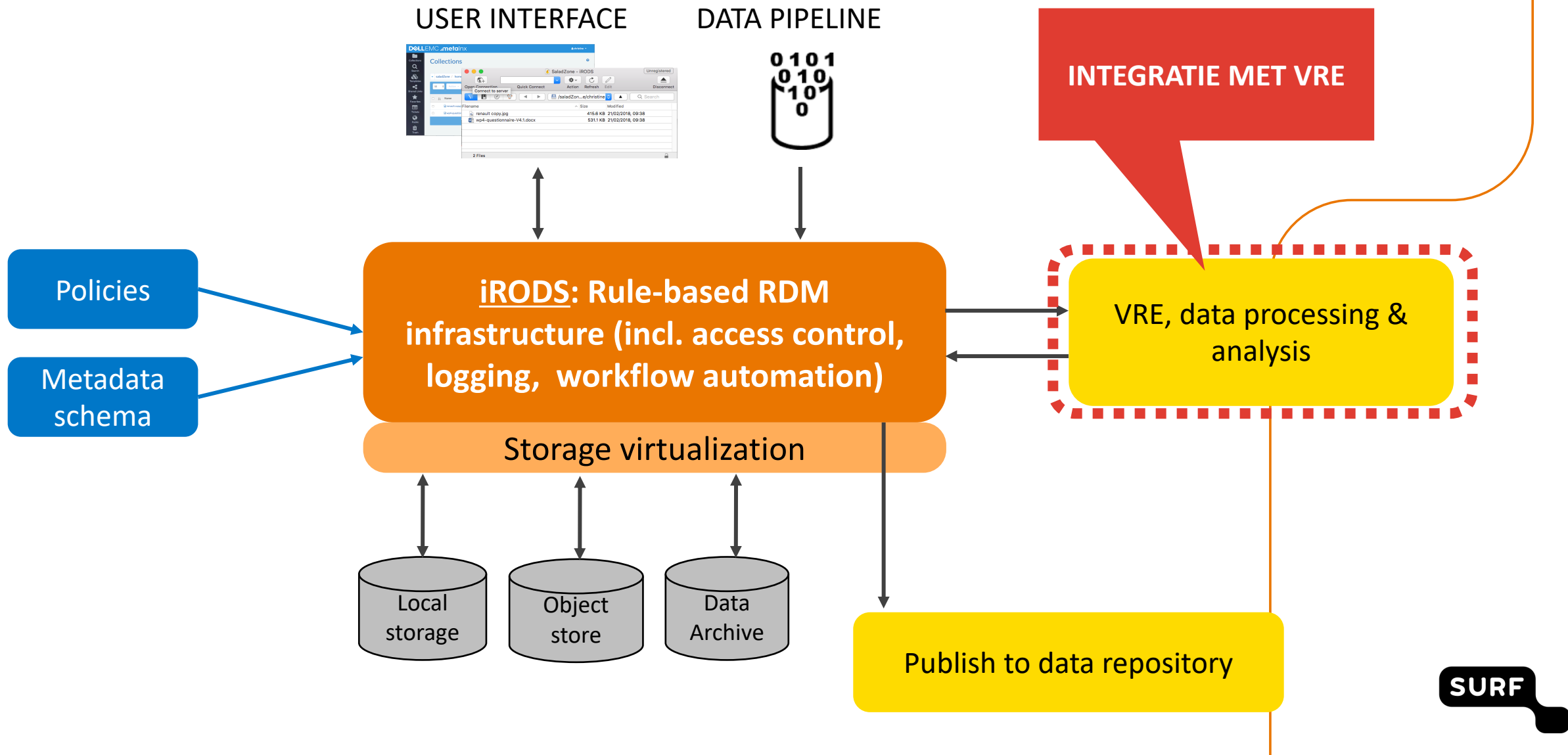
Volgende modules (nog te prioriteren)



Volgende modules (nog te prioriteren)



Volgende modules (nog te prioriteren)



Samenwerking binnen RDM Expertise Centrum (“RDM-TEC”)

- Overleg orgaan met vertegenwoordiging van (bijna) alle Nederlandse universiteiten
- Afstemming ‘vraag’ en ‘aanbod’:
 - Vraag: waar ligt de grootste gezamenlijke behoefte?
 - Aanbod: afstemming roadmap dienstontwikkeling producenten (UU, RUG, SURF) ⇒ prioritering RDM modules
- Focus op iRODS
- Pilot, review eind 2019
- Concrete activiteit: organiseren van pilots met iRODS & YODA binnen vier verschillende instellingen.

RESEARCH DATA MANAGEMENT & ACCESS OF RESEARCH WORKSPACES



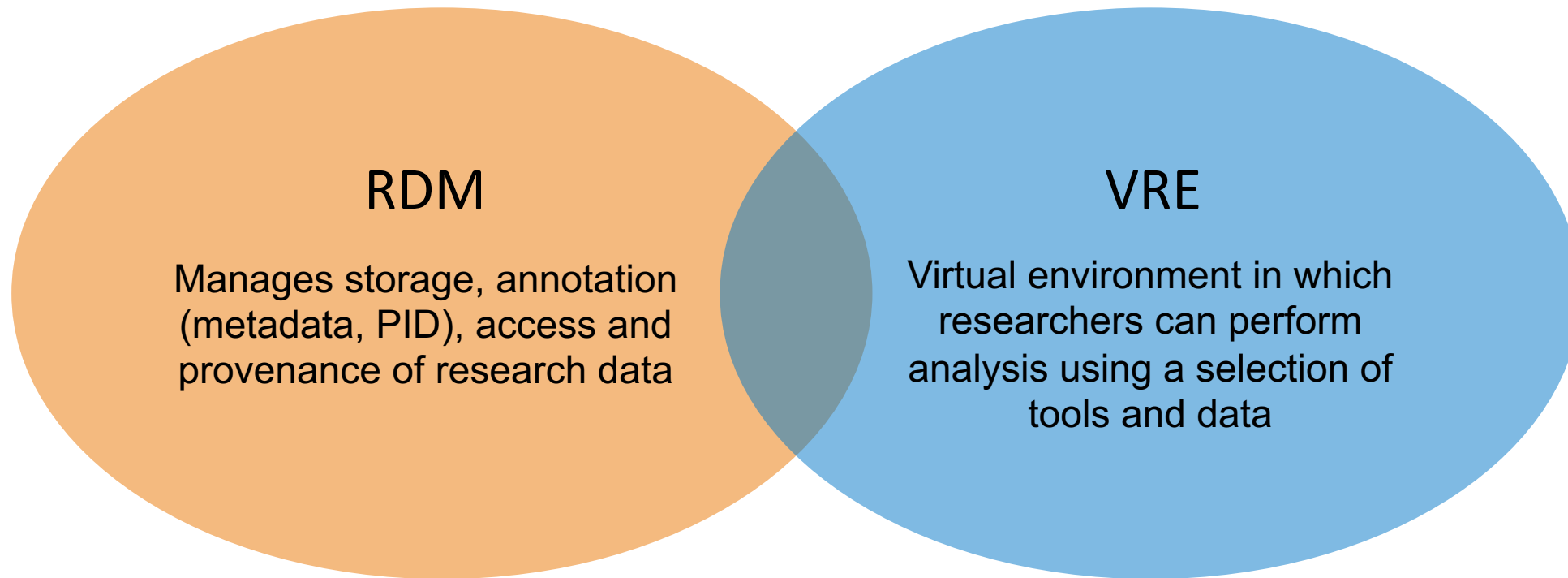
SURF

Who needs RDM when you have a VRE?

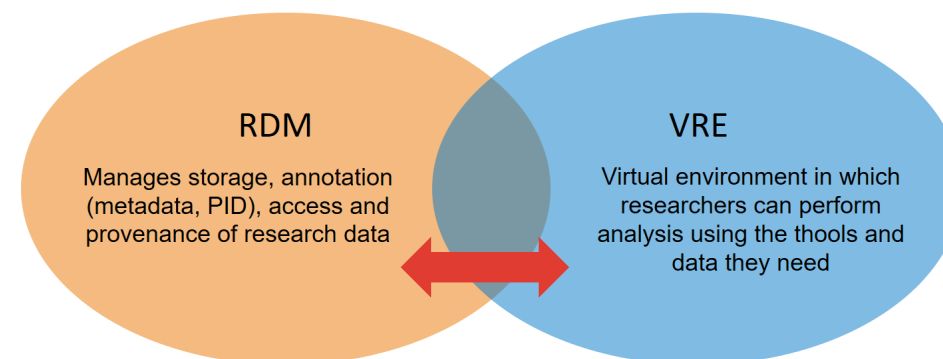
Iedereen! Goed beheer van de data zorgt voor:

- Vindbaarheid van data (metadata en meer)
- “Track-and-trace” (provenance) van data van experiment tot analyse tot publicatie/archivering
- Data (automatisch) beschikbaar maken in de virtuele werkomgeving
- Data publiceren in een data repository vanuit de werkomgeving
- Automatisering van data opslag en de bijbehorende trade-offs zoals:
 - Data dicht bij compute nodes voor snelle bewerking; vs.
 - Data op tape voor kosteneffectieve lange-termijn opslag.
- Beheren van metadata volgens community standaarden
- Toegangsbeheer

Integratie tussen RDM & VRE



Integratie tussen RDM & VRE



TO DO: Add analysis possible connections:
API's, clients, bridges, extension.

TO DO: Connect with the need for solid AAI
solution underpinning the integration