



SURF

Data Management with SURF

Alice Stuart-Lee
Claudio Cacciari
Jorik van Kemenade
Maithili Kalamkar-Stam
Steven Voges

25-01-2024



| What are we covering

- 09:00 – 09:45: Data management with SURF
- 10:00 – 11:00: Data creation and analysis with Research Cloud and Research Drive
- 11:15 – 13:00: Data management in Yoda



| What are we covering

- What is Research Data Management?
- The research data lifecycle.
- What is Metadata?
- The data management landscape, according to SURF.

Research Data:

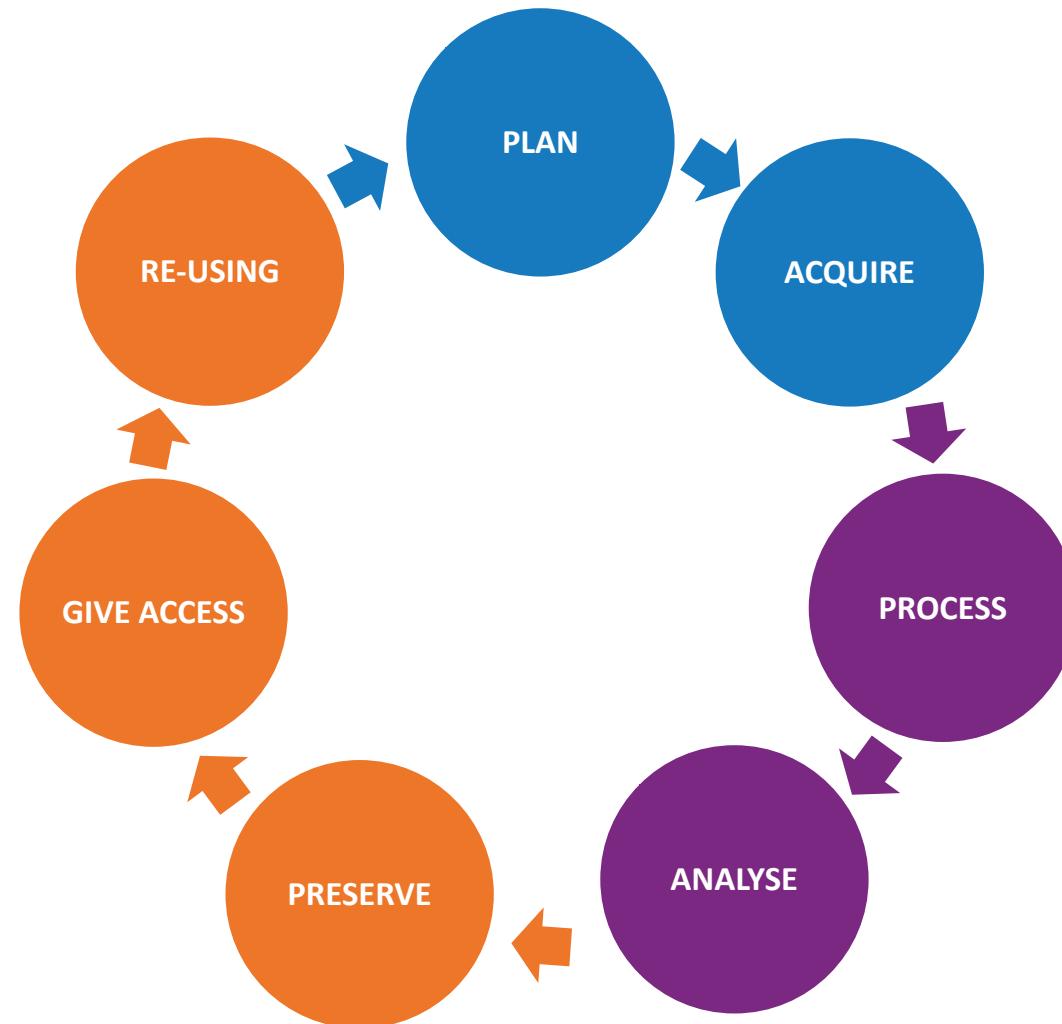
“materials generated or collected during the course of conducting research”,

by The National Endowment for the Humanities.

Data Management:

“Actions that contribute to effective storage, preservation and reuse of data and documentation throughout the research lifecycle.”

Research Data Life Cycle – the holy grail of Research Data Management



Why data management?



The data sharing advantage in astrophysics

S. B. F. Dorch, T. M. Drachen, O. Ellegaard

We present here evidence for the existence of a citation advantage within astrophysics for papers that link to data. Using simple measures based on publication data from NASA Astrophysics Data System we find a citation advantage for papers with links to data receiving on the average significantly more citations per paper than papers without links to data. Furthermore, using INSPEC and Web of Science databases we investigate whether either papers of an experimental or theoretical nature display different citation behavior.

Comments: 4 pages, 2 figures, Conference proceedings of Focus Meeting 3 on Scholarly Publication in Astronomy, IAU GA 2015, Honolulu

Subjects: **Instrumentation and Methods for Astrophysics (astro-ph.IM); Digital Libraries (cs.DL)**

Cite as: [arXiv:1511.02512 \[astro-ph.IM\]](https://arxiv.org/abs/1511.02512)
(or [arXiv:1511.02512v1 \[astro-ph.IM\]](https://arxiv.org/abs/1511.02512v1) for this version)
<https://doi.org/10.48550/arXiv.1511.02512>

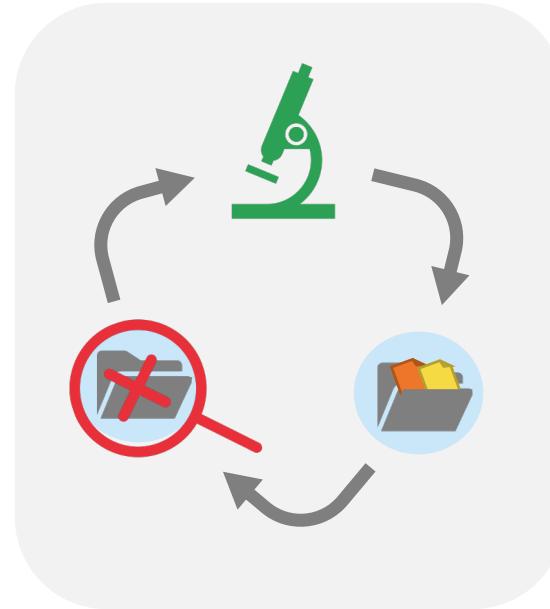
Consequences of bad research data management



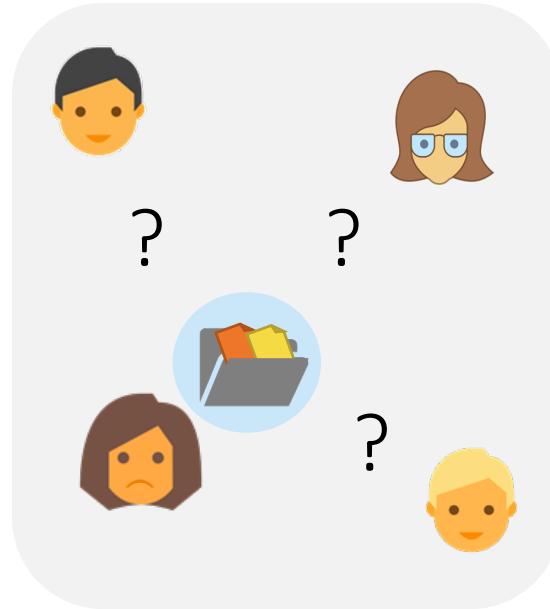
No cost-effective data storage



Data gets lost by disaster or loss of context

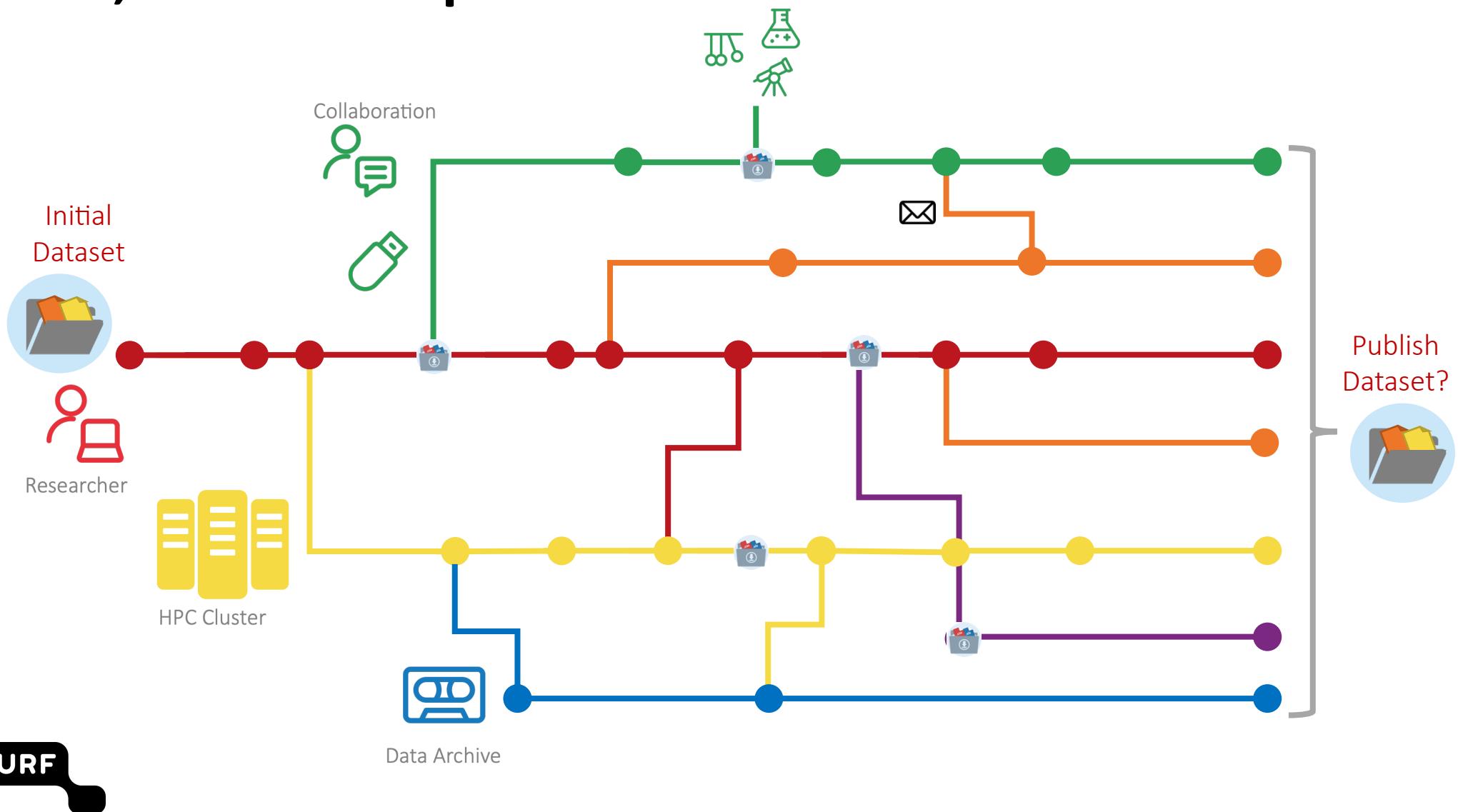


Redoing experiments and no interoperability.



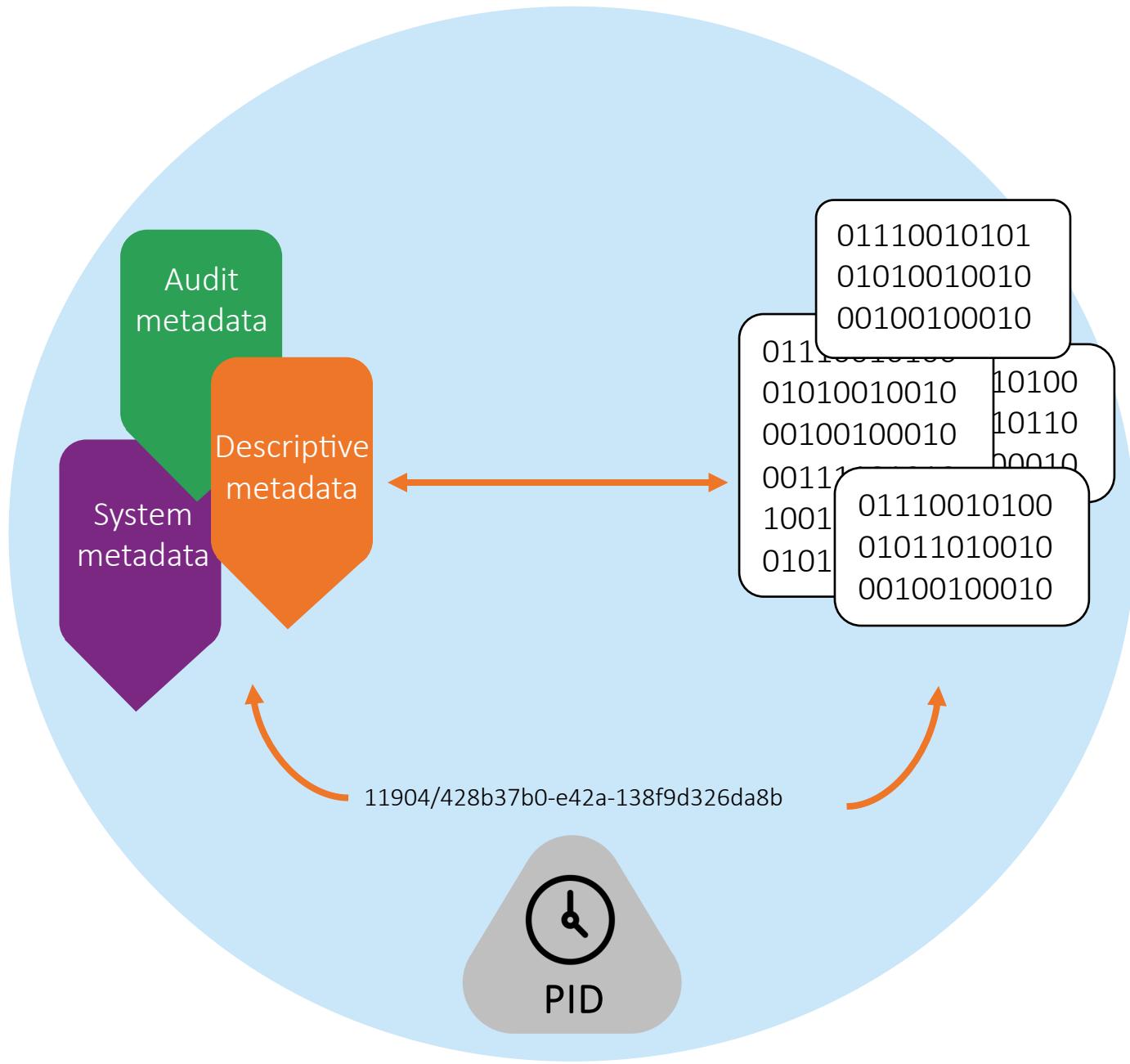
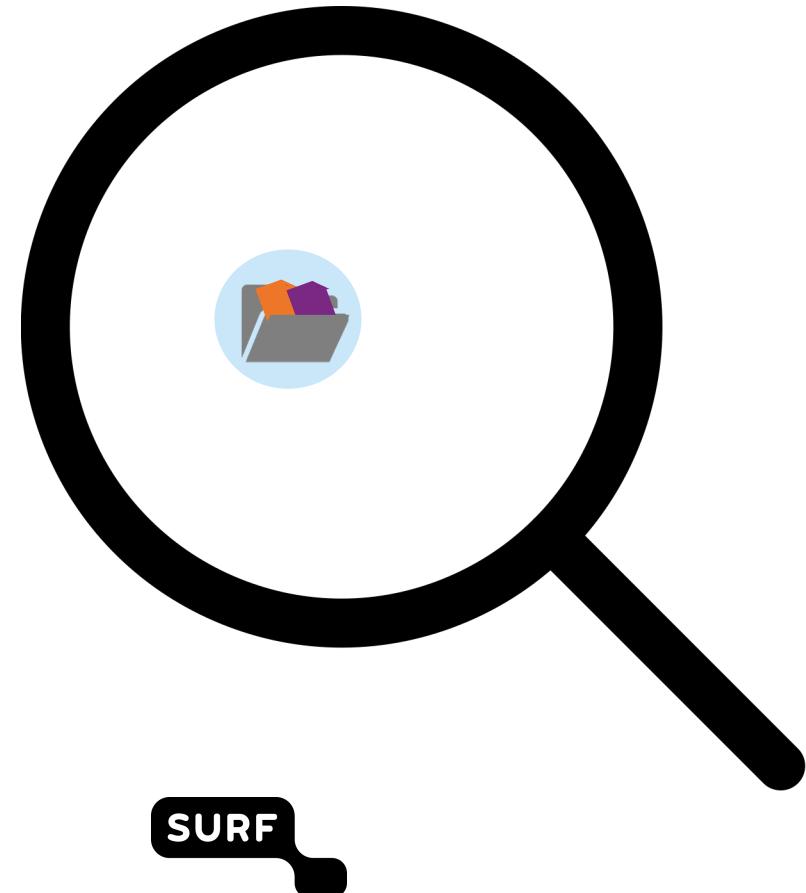
Not able or fear to share / publish data.

Data, what's the problem?





Data, what is it?



| Metadata – What is the problem?

“A package of cookies contains more metadata than most research data packages”



| Metadata– Help others understand your products



Visual description

Text description

Organisation

Metadata – Help others understand your products



Interoperability

Persistent Identifier

Technical information

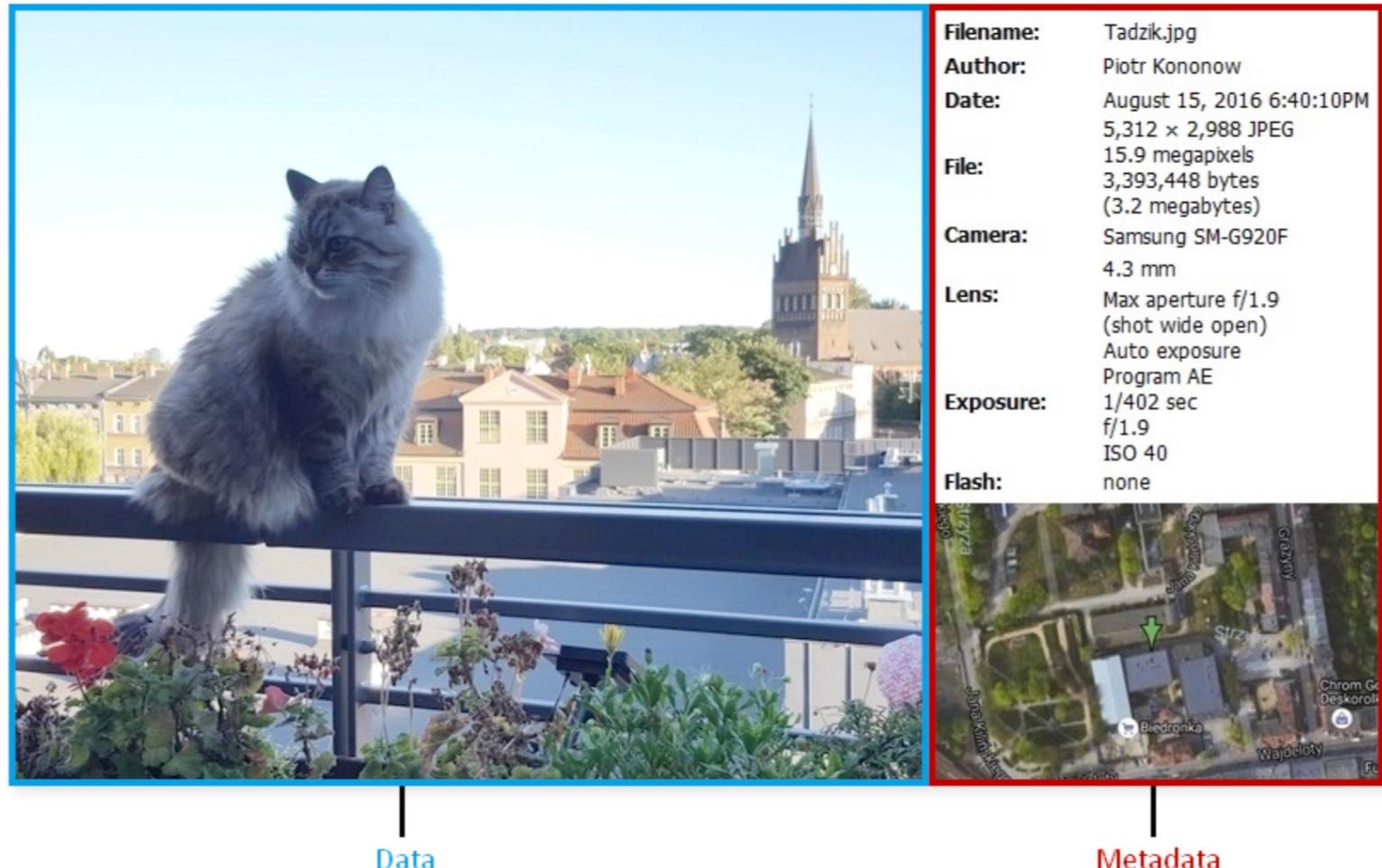
| Metadata – Sell your research products

“Would you buy
cookies without
metadata?”

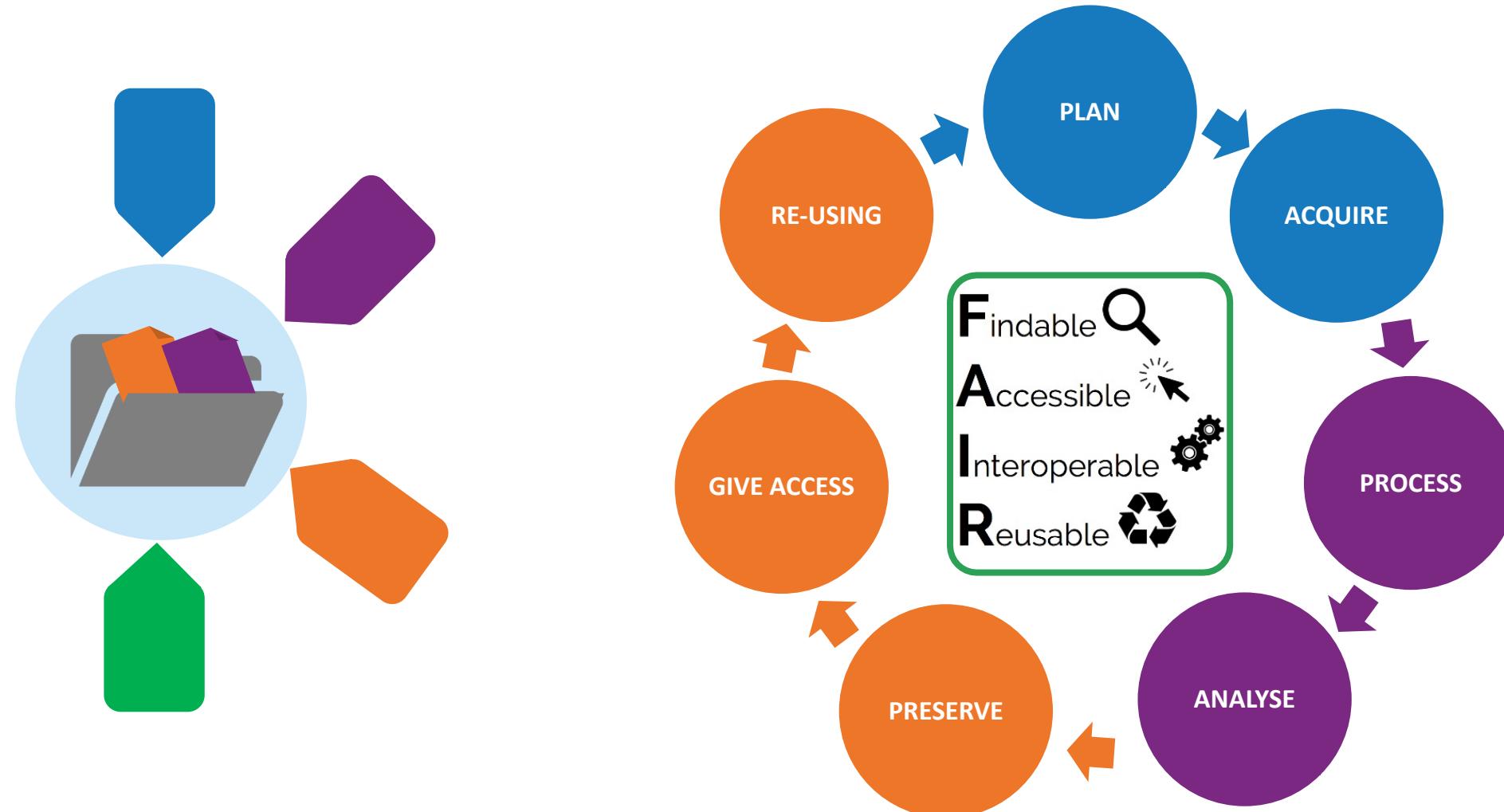


Why use metadata?

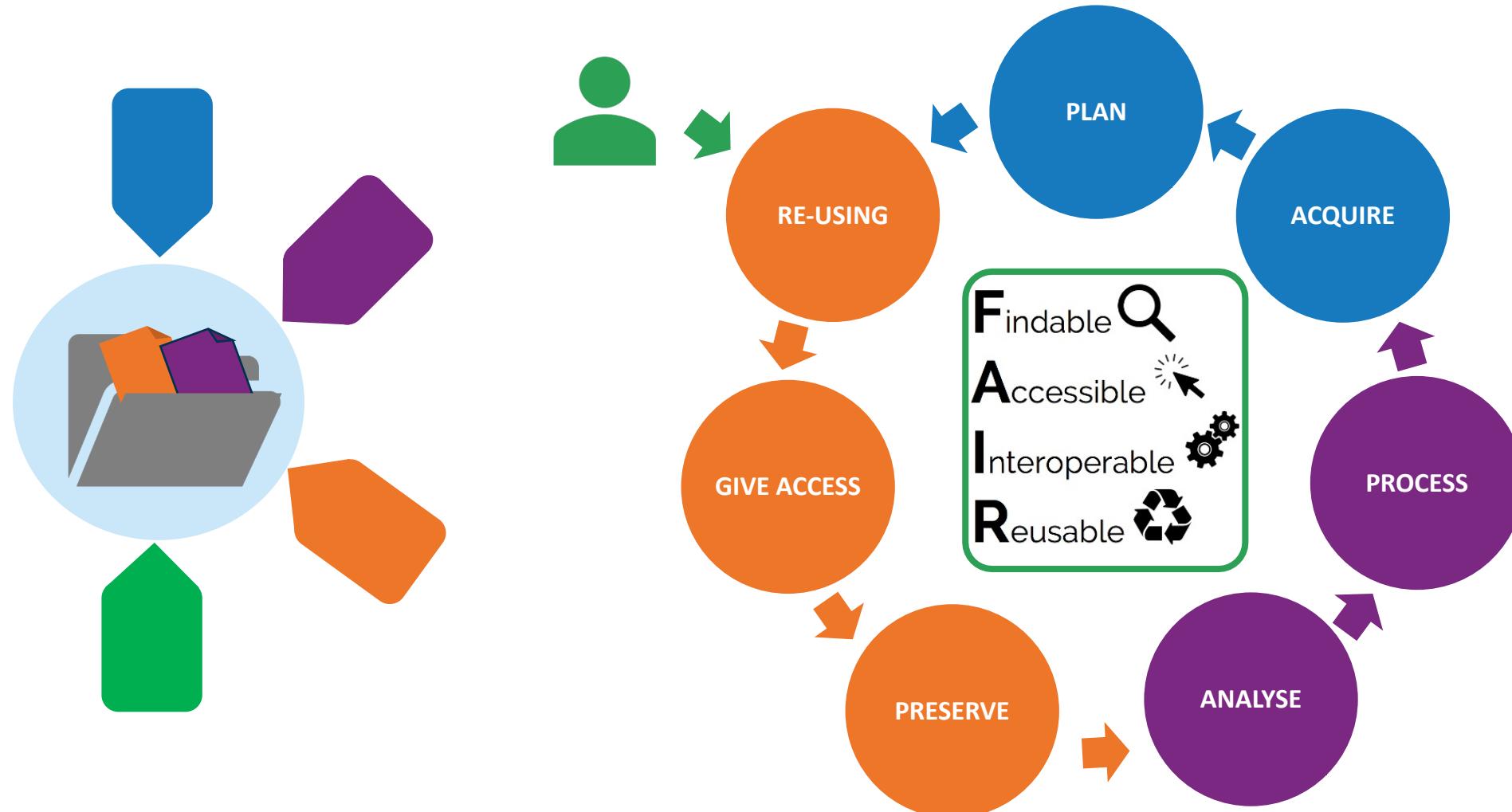
- Facilitate data discovery.
- Help users determine the applicability of the data.
- Enable interpretation and reuse of data.
- Allow any limitations to be understood.
- Clarify ownership and restrictions on reuse.



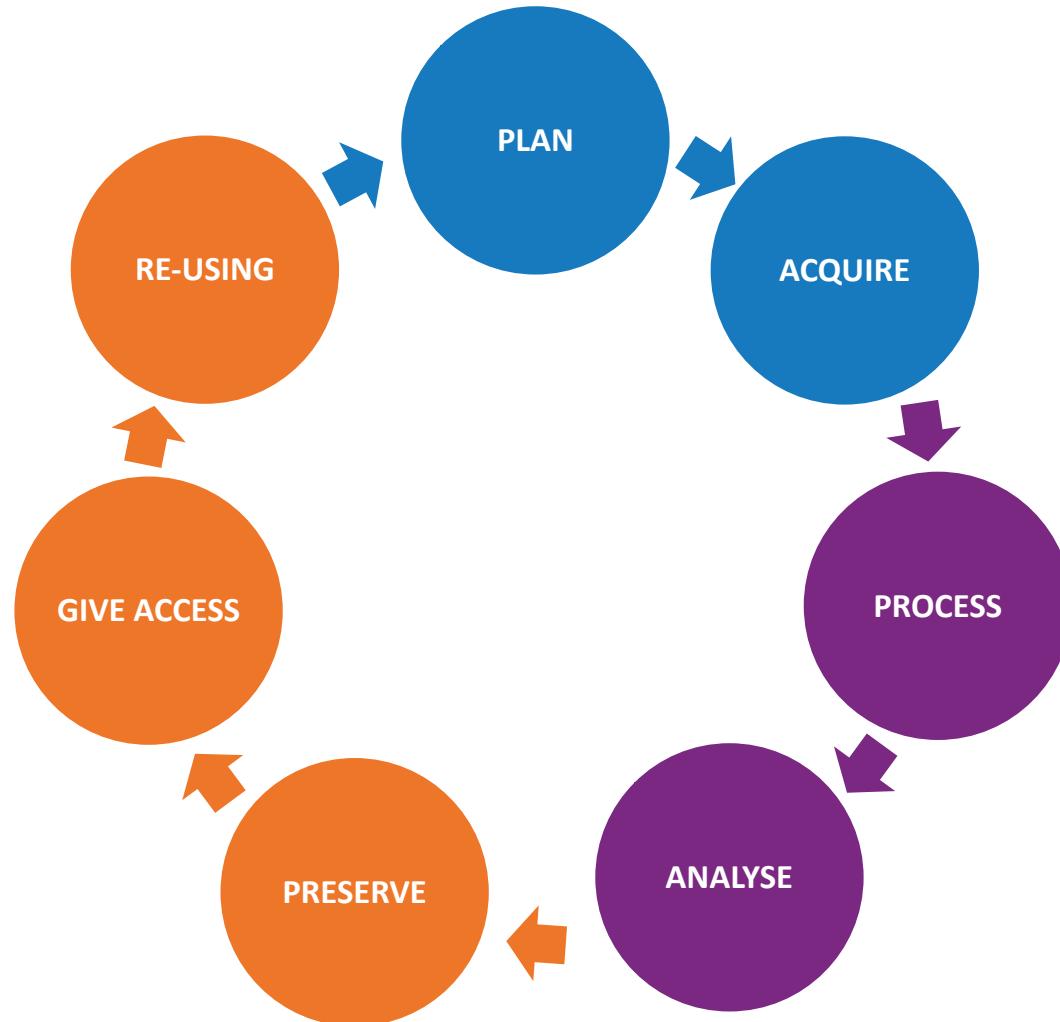
Research Data Lifecycle described by metadata



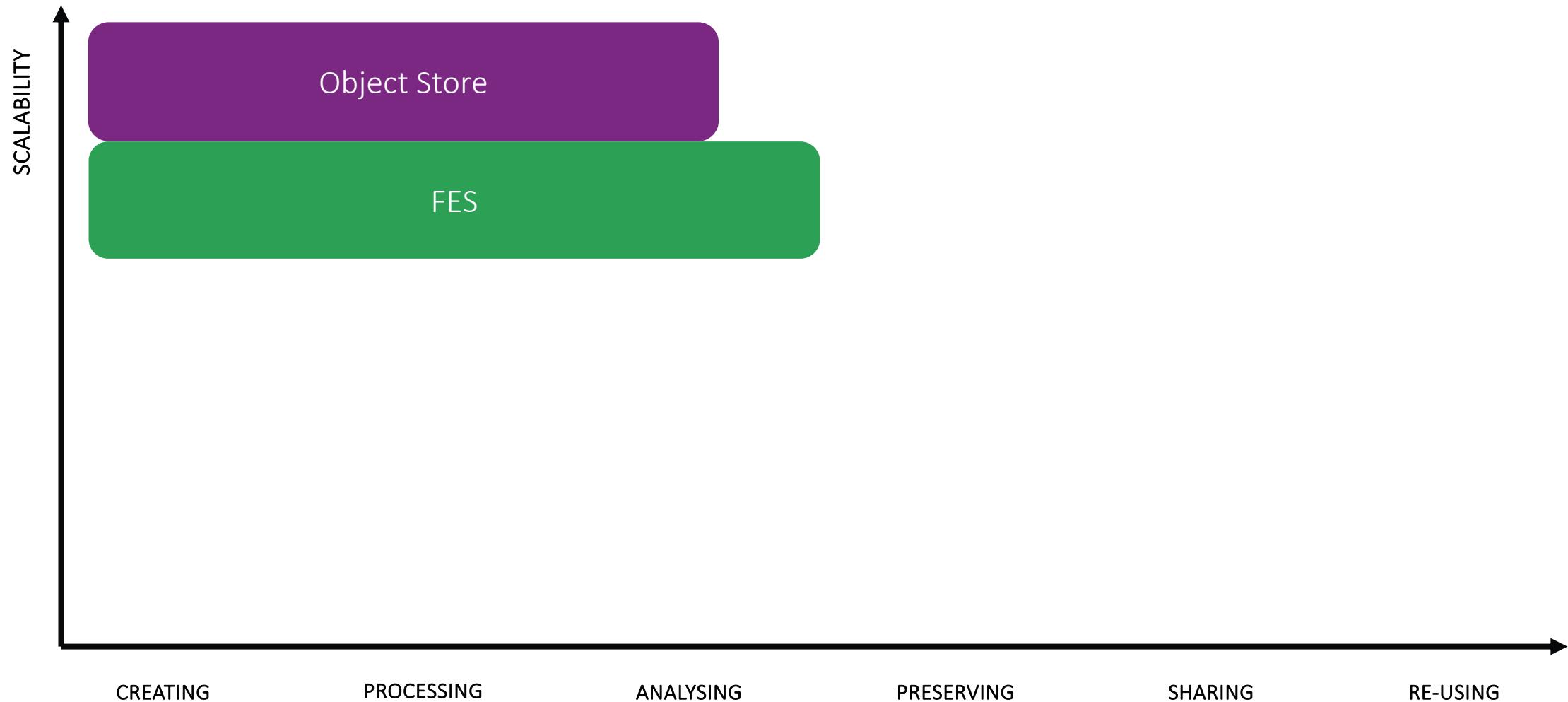
Research Data Lifecycle as a re-user



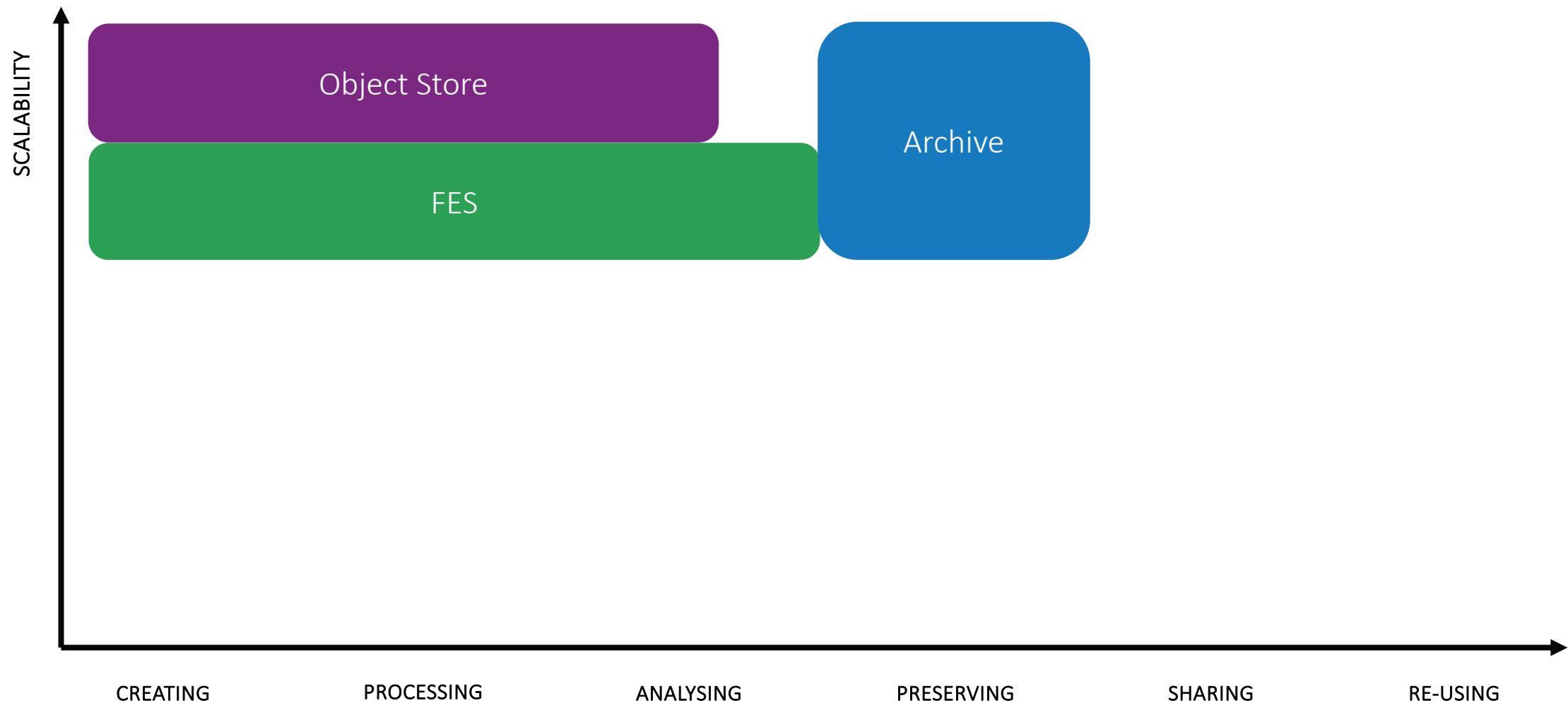
Research Data Life Cycle – the holy grail of Research Data Management



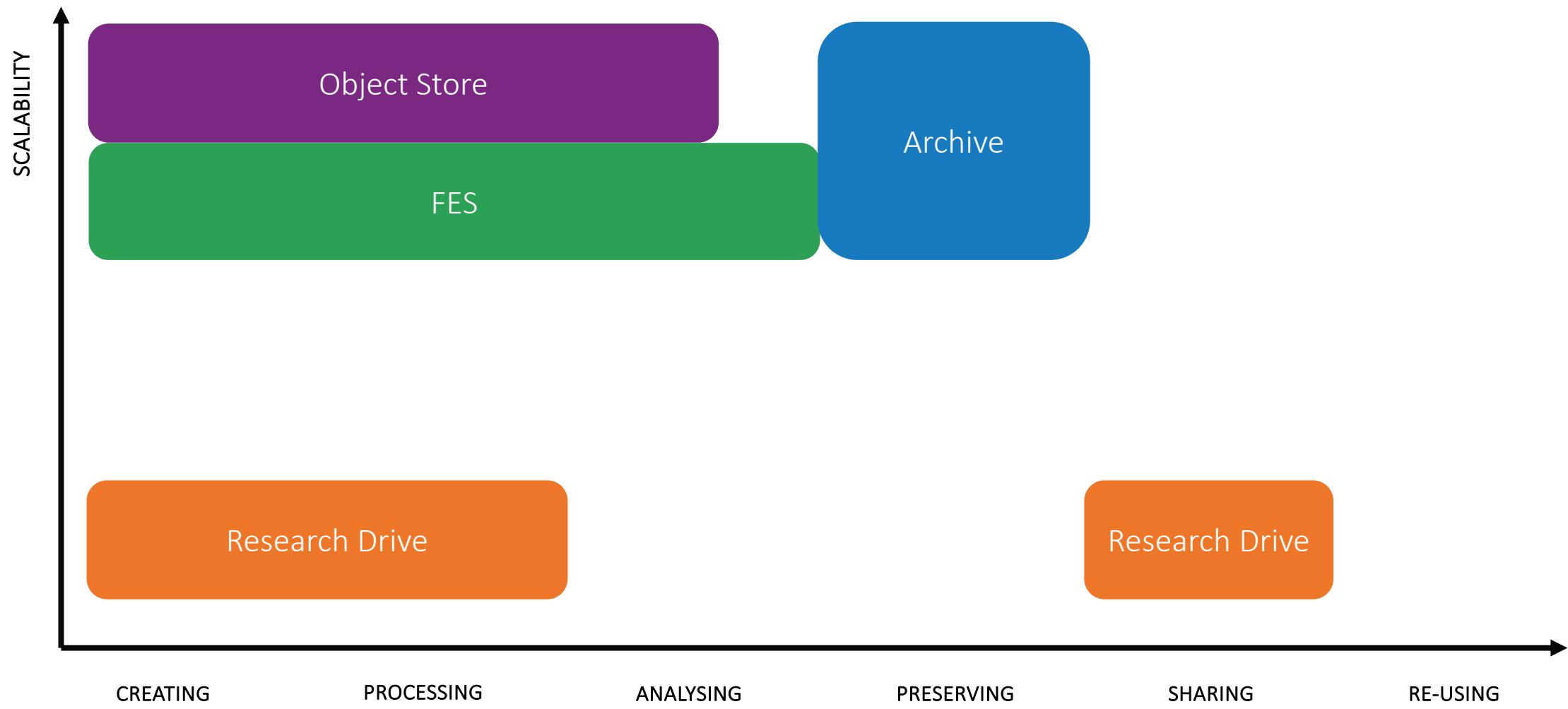
Services at SURF Data Services



Services at SURF Data Services



Services at SURF Data Services



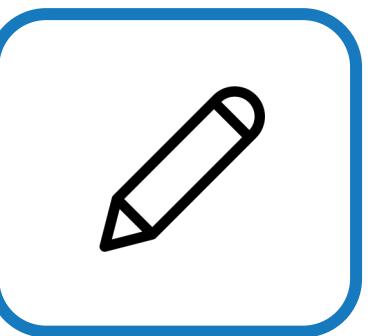
| What is Research Drive



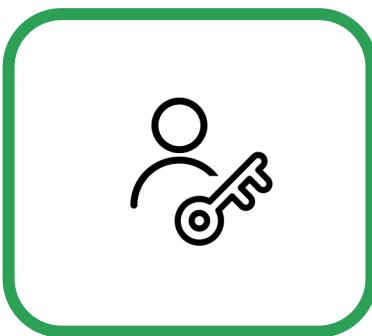
Cloud
environment



Sharing
files & folders



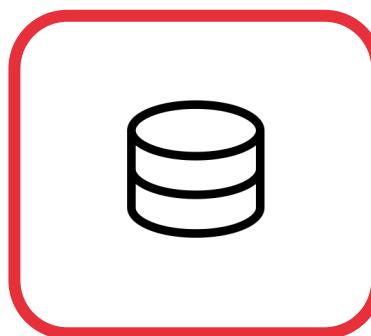
Collaborative
editing



Role Based
Data
Management

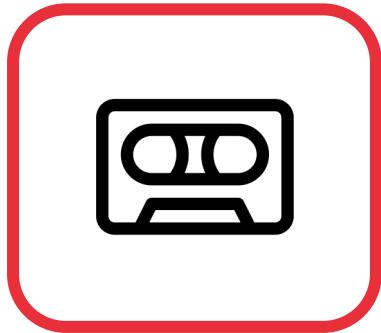


Access for
Guests



Unlimited
storage

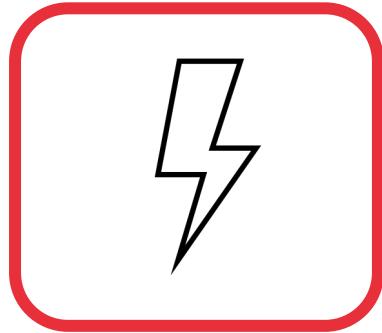
| What is Research Drive not



Long-term archiving



Data
publishing



Super Fast
Data Analysis

Store and access data

Web Brower
Interface



OwnCloud
desktop client



WebDav
client



- Data stays in cloud
- Access and modification of data via build-in applications or download

- Local copy of data is created on machine
- Access and modification of data offline possible
- Automatic resyncing

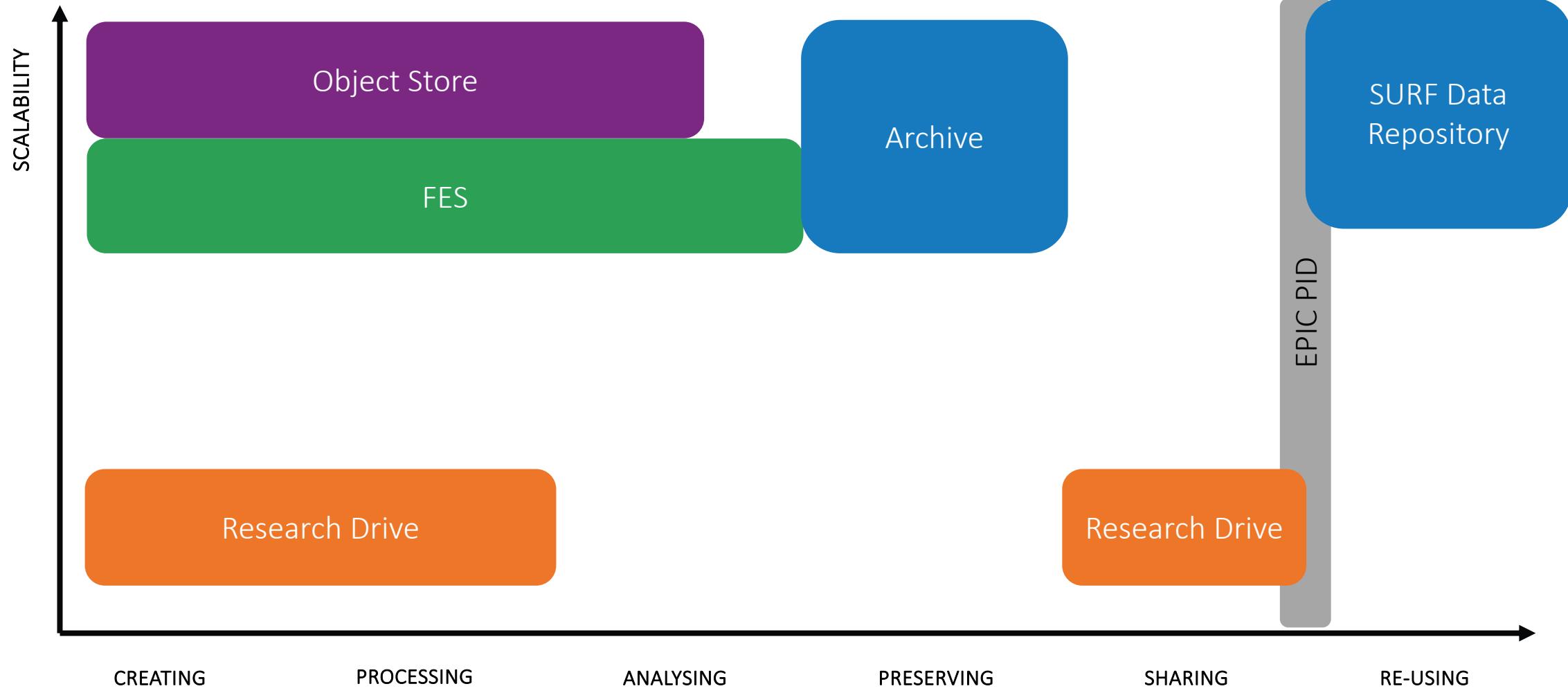
- Data stays in cloud but is visible as file system in computer
- Access and modification of data via local applications

- No access for offline users
- Limited amount of build in applications
- Modified data needs upload again

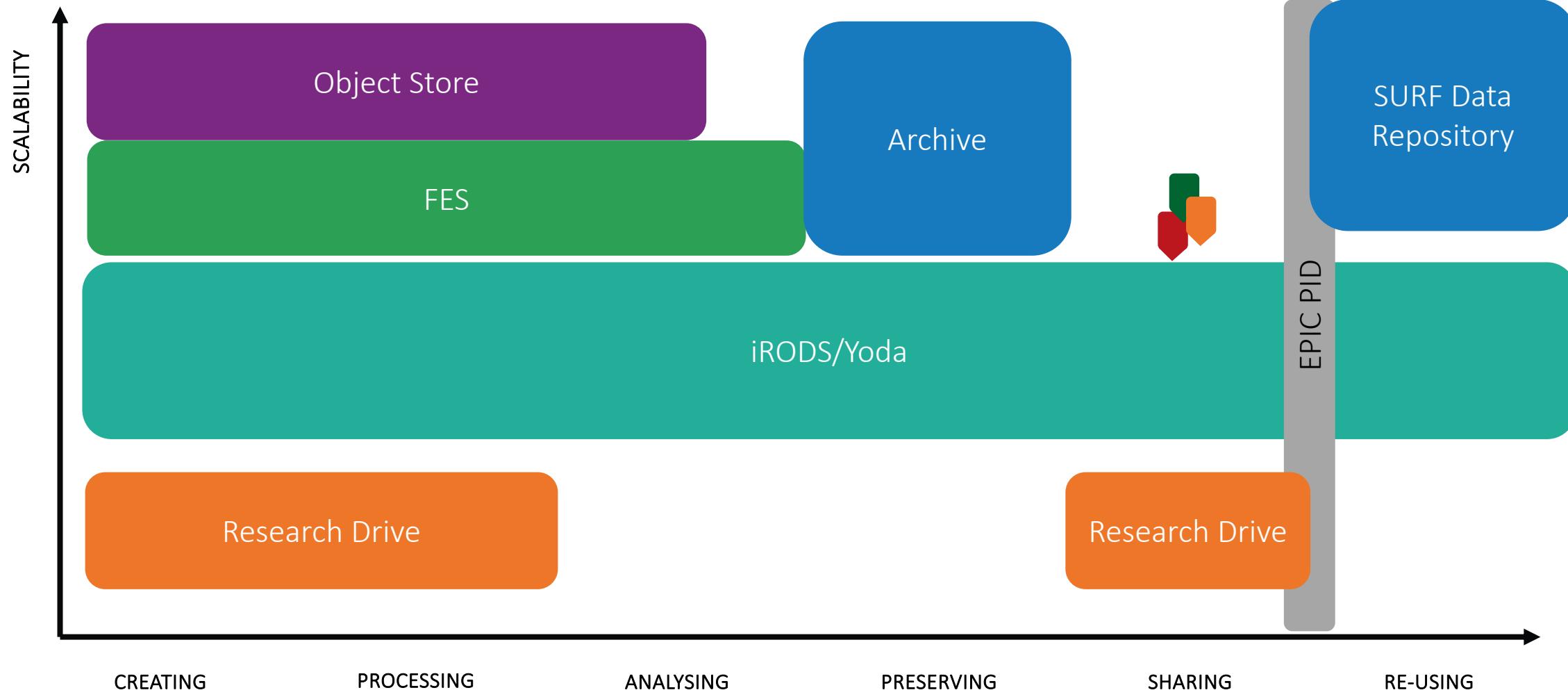
- “Out-of sync” data is possible
- Large amount of data should not be synchronized
- Local copy of data is created on machine

- No access for offline users
- Limited capacity when it comes to many files

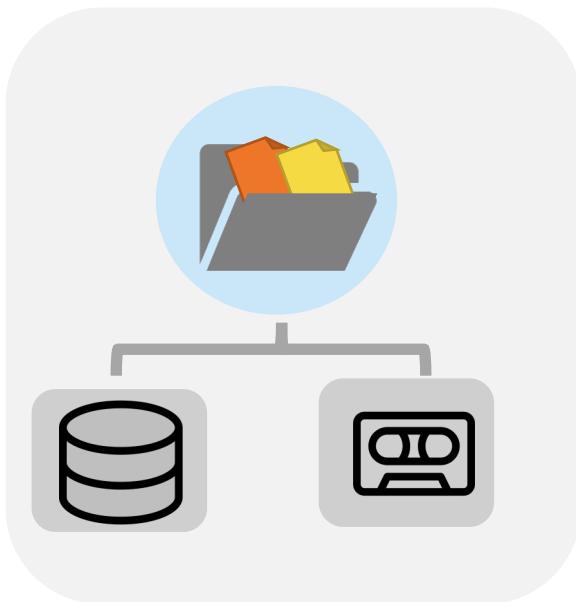
Services at SURF Data Services



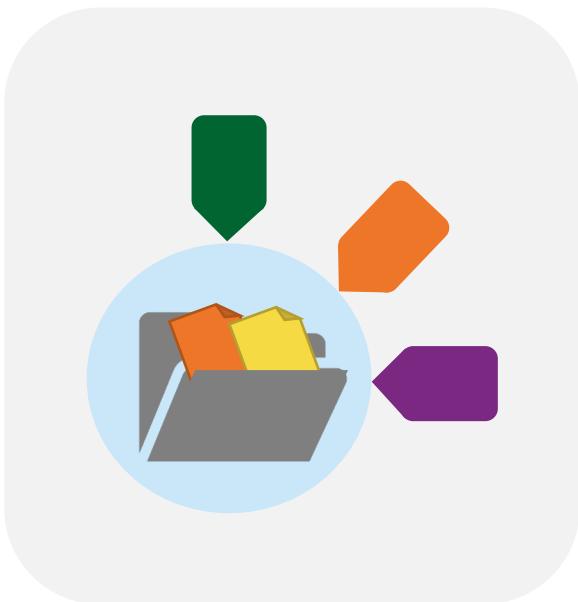
Services at SURF Data Services



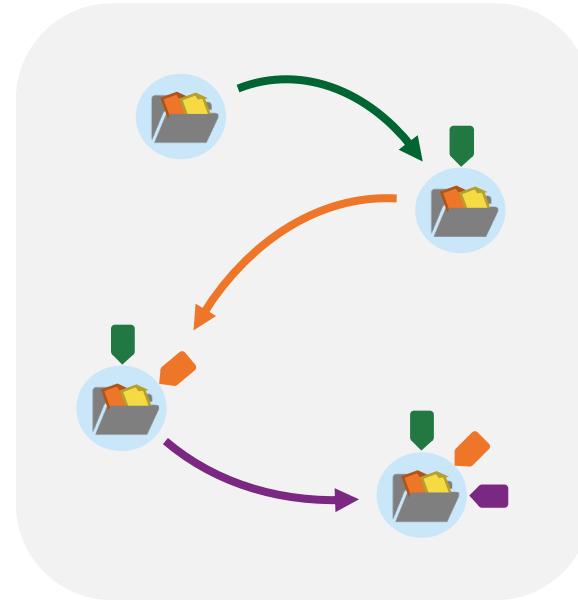
| What is iRODS?



Unified storage of
disk and tape



Metadata for data
discovery

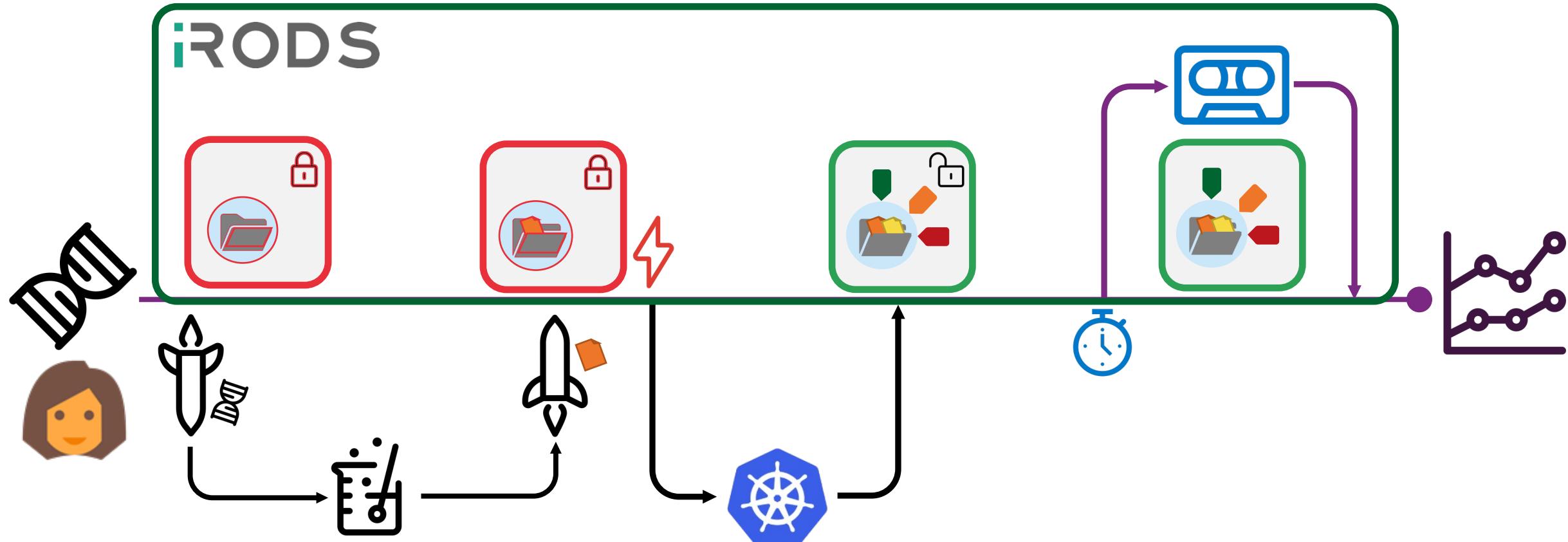


Secure collaboration
and auditing

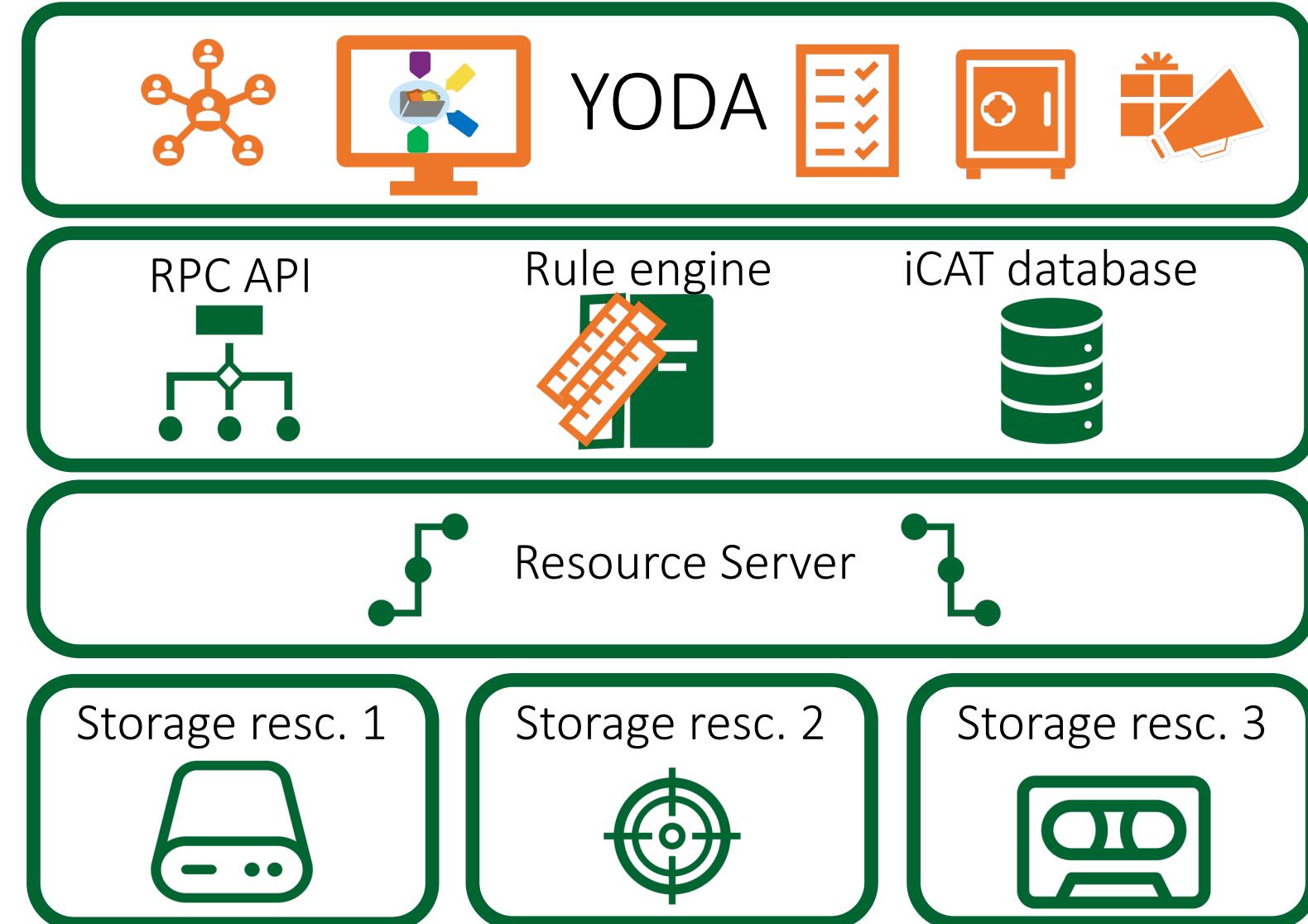


Rule engine to
automate policies

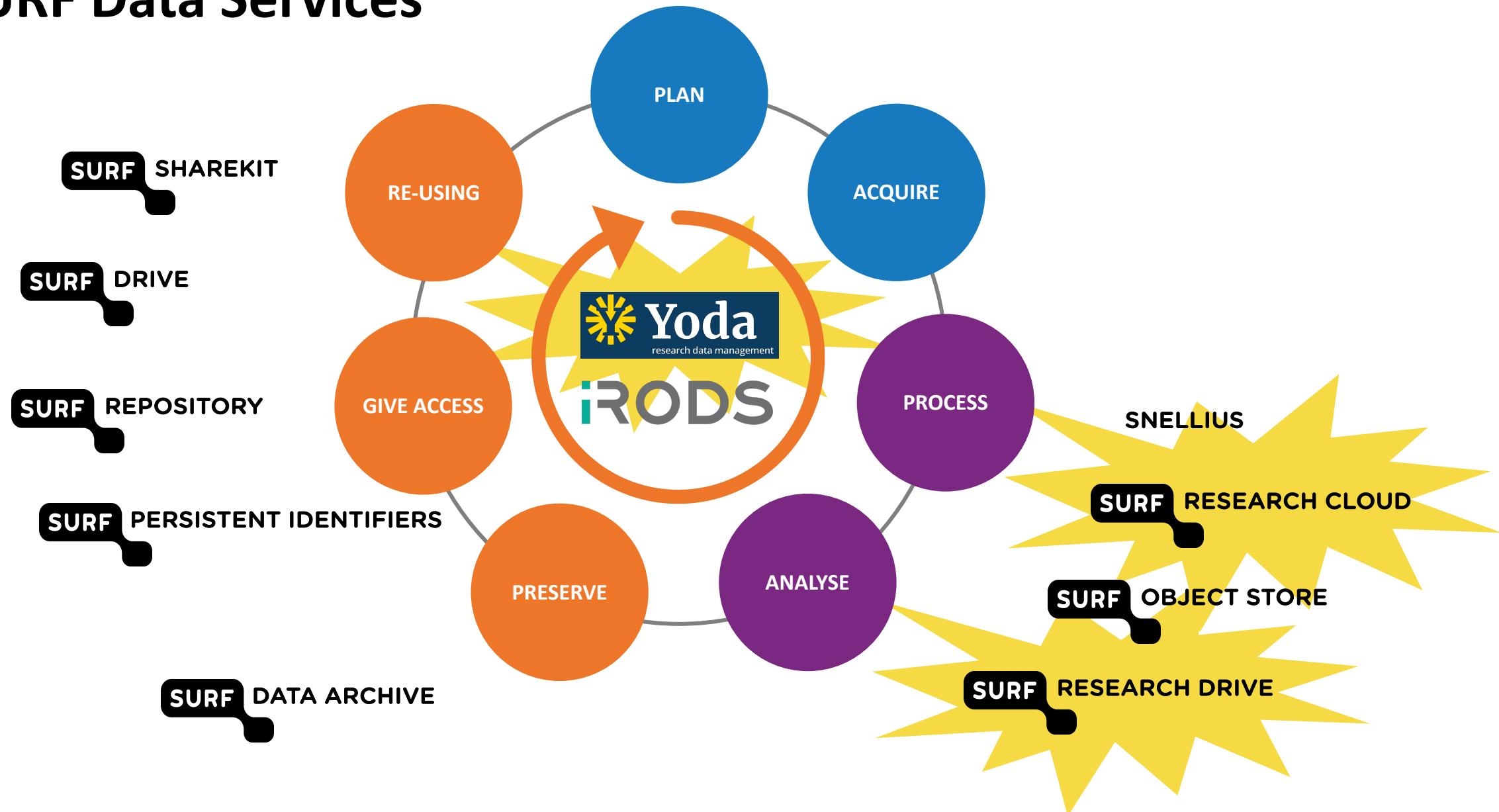
iRODS User journey: DNA sequencing (real use case)

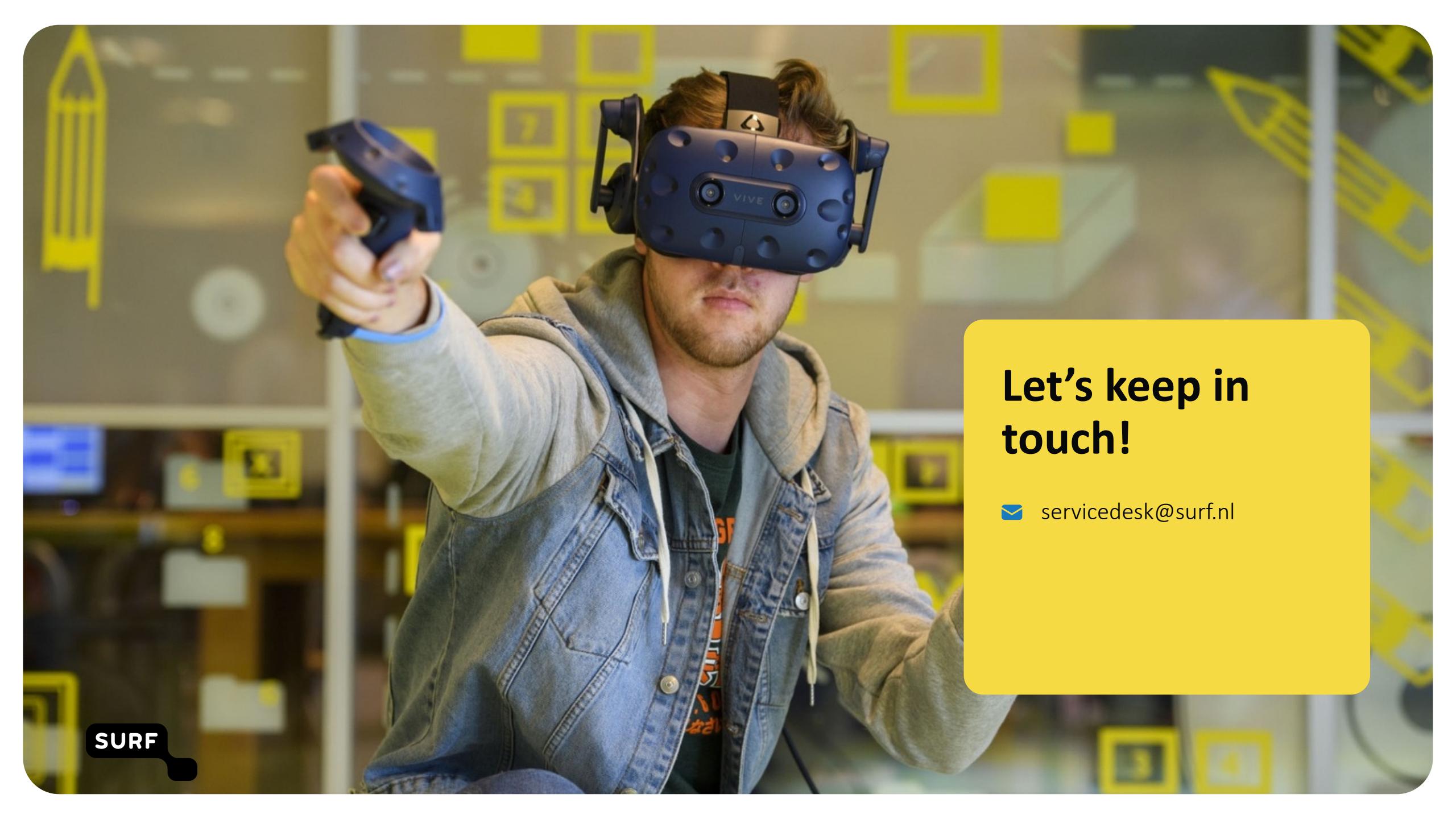


| Yoda – an opinionated use of iRODS



SURF Data Services





SURF

**Let's keep in
touch!**

✉️ servicedesk@surf.nl

SURF