ASTRON
Netherlands Institute for Radio Astronomy

ASTRON Science Data Centre (SDC):
Science analysis platform and datalake

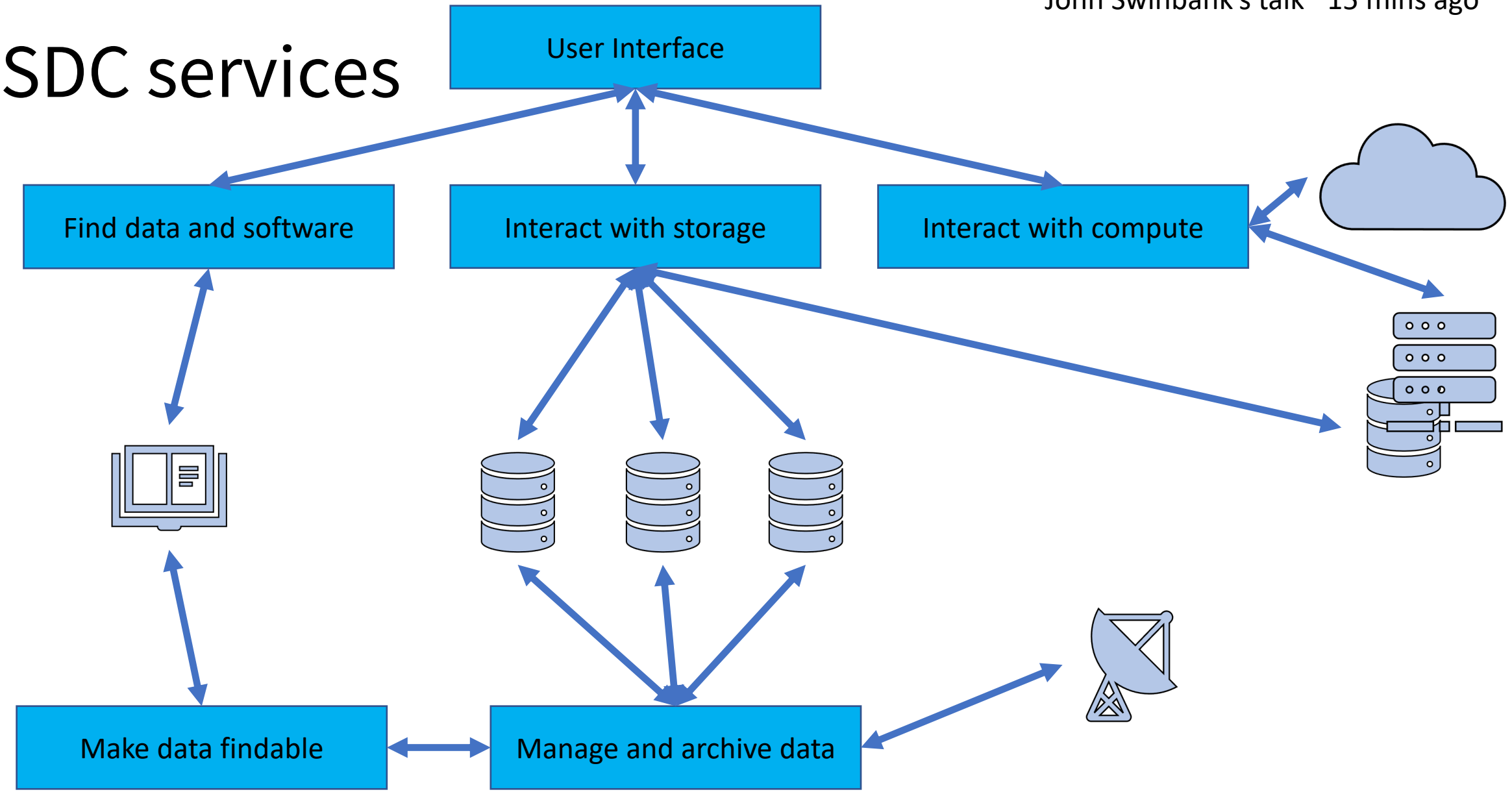
Yan Grange, *John Swinbank*, *Zheng Meyer-Zhao*,
Nico Vermaas, *Mattia Mancini*, *Hanno Holties*

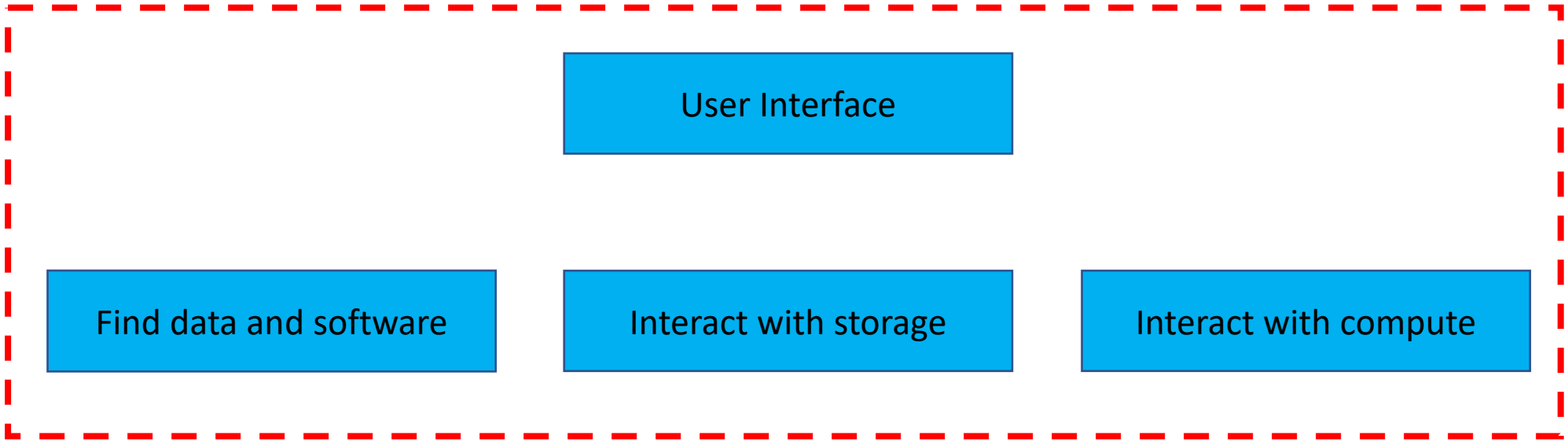
ASTRON
Netherlands Institute for Radio Astronomy

ESCAPE
European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

Assuming you remember
John Swinbank's talk ~15 mins ago

SDC services





Science Analysis Platform

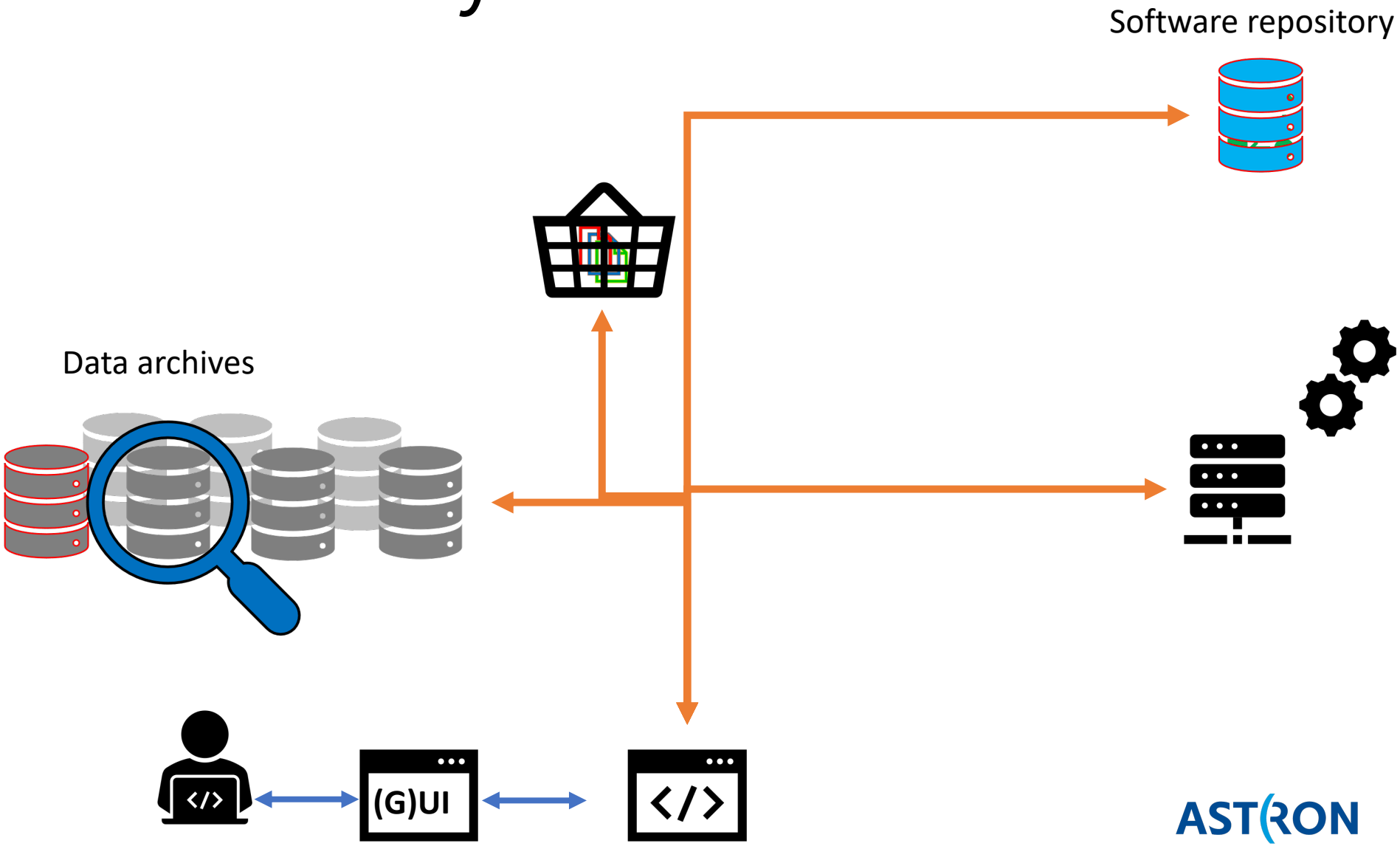
Make data findable

Manage and archive data

ASTRON
Netherlands Institute for Radio Astronomy

ESCAPE
European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

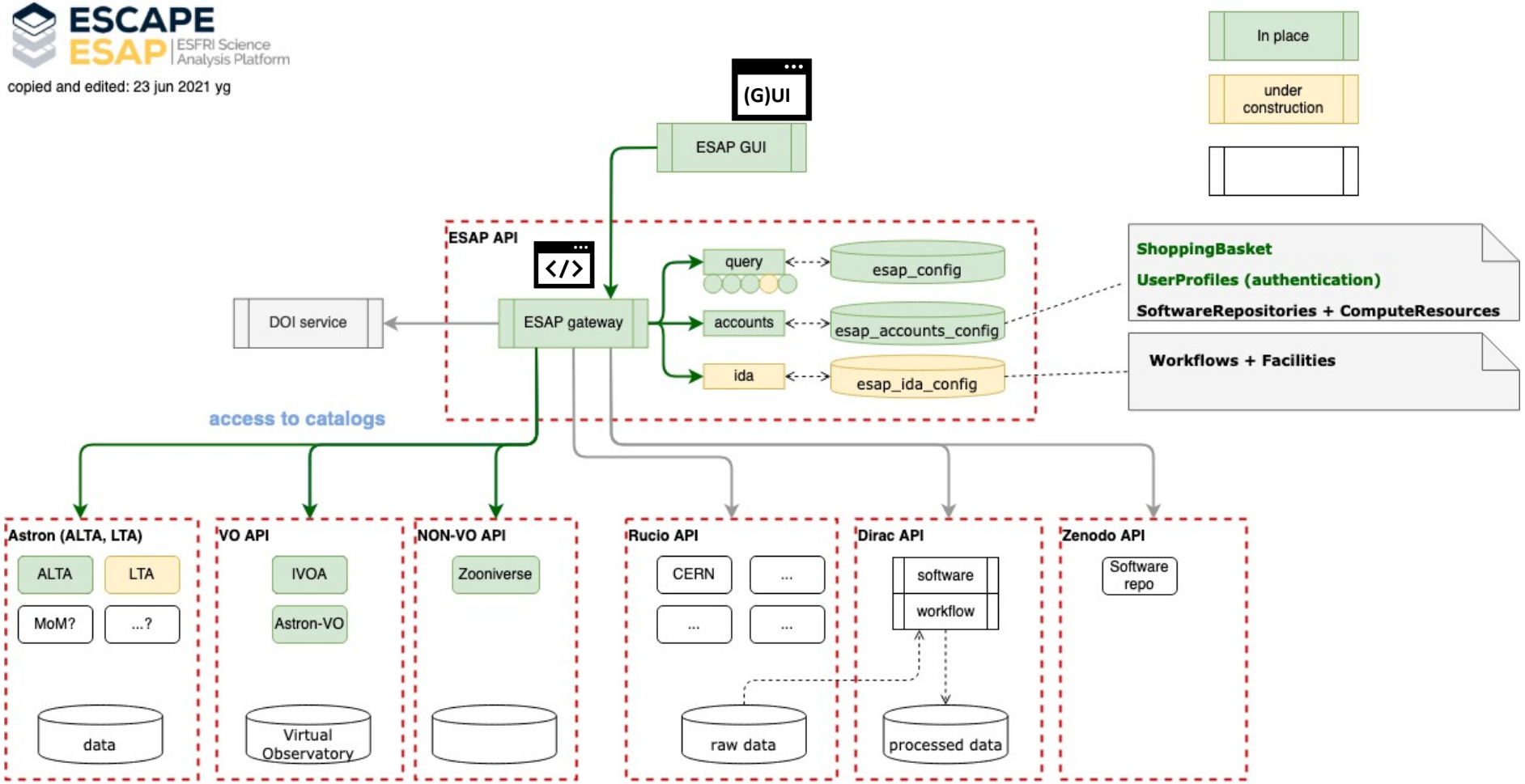
Science Analysis Platform



ASTRON
Netherlands Institute for Radio Astronomy

ESCAPE
European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

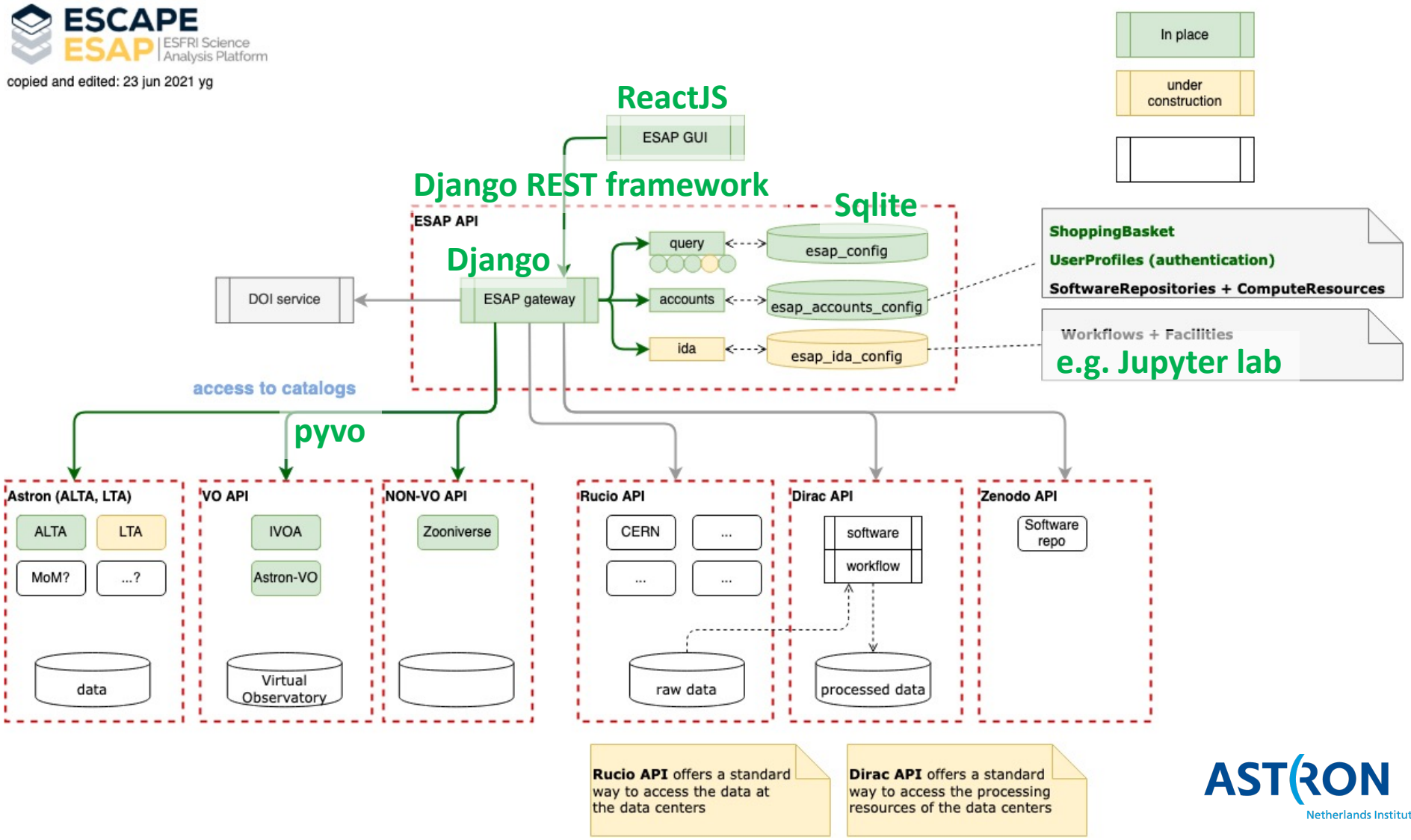
Science analysis platform architecture



ASTRON
Netherlands Institute for Radio Astronomy

ESCAPE
European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

Science analysis platform architecture



ASTRON
Netherlands Institute for Radio Astronomy

ESCAPE
European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

Science platform

ESAP

ESAP API Gateway - Query

"data model API endpoints"

API resources

Function	URL
Archives	/esap-api/query/archives
Datasets	/esap-api/query/datasets
Catalogs	/esap-api/query/catalogs
Parameter Mapping	/esap-api/query/parameter-mapping

click in browser

curl

python requests

javascript promise

etc...

ASTRON - version 23 mar 2021 - 10:00

ASTRON
Netherlands Institute for Radio Astronomy

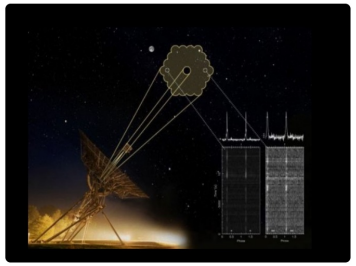
ESCAPE
European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

Science platform

ESAP Archives Query Rucio Interactive Analysis SAMP Login

archives table

WSRT-Apertif



Apertif Surveys

Data from the Apertif surveys include imaging and time-domain data. The time-domain products consist of high-time resolution filterbank data in the PSRFITS standard. The imaging data products include the raw observations in the measurement set (MS) standard format. In addition, processed data products are available, including calibration tables, calibrated visibilities, multi-frequency synthesis continuum images, polarization images and

ASTRON VO

ASTRON Virtual Observatory

The Virtual Observatory defines a set of standards that can be used to download astronomical data. The ASTRON VO contains several image surveys, which are images in the FITS format. Since the VO is currenty under development, more data types will be available in the future.

Visit ASTRON VO Archives

LOFAR-LTA

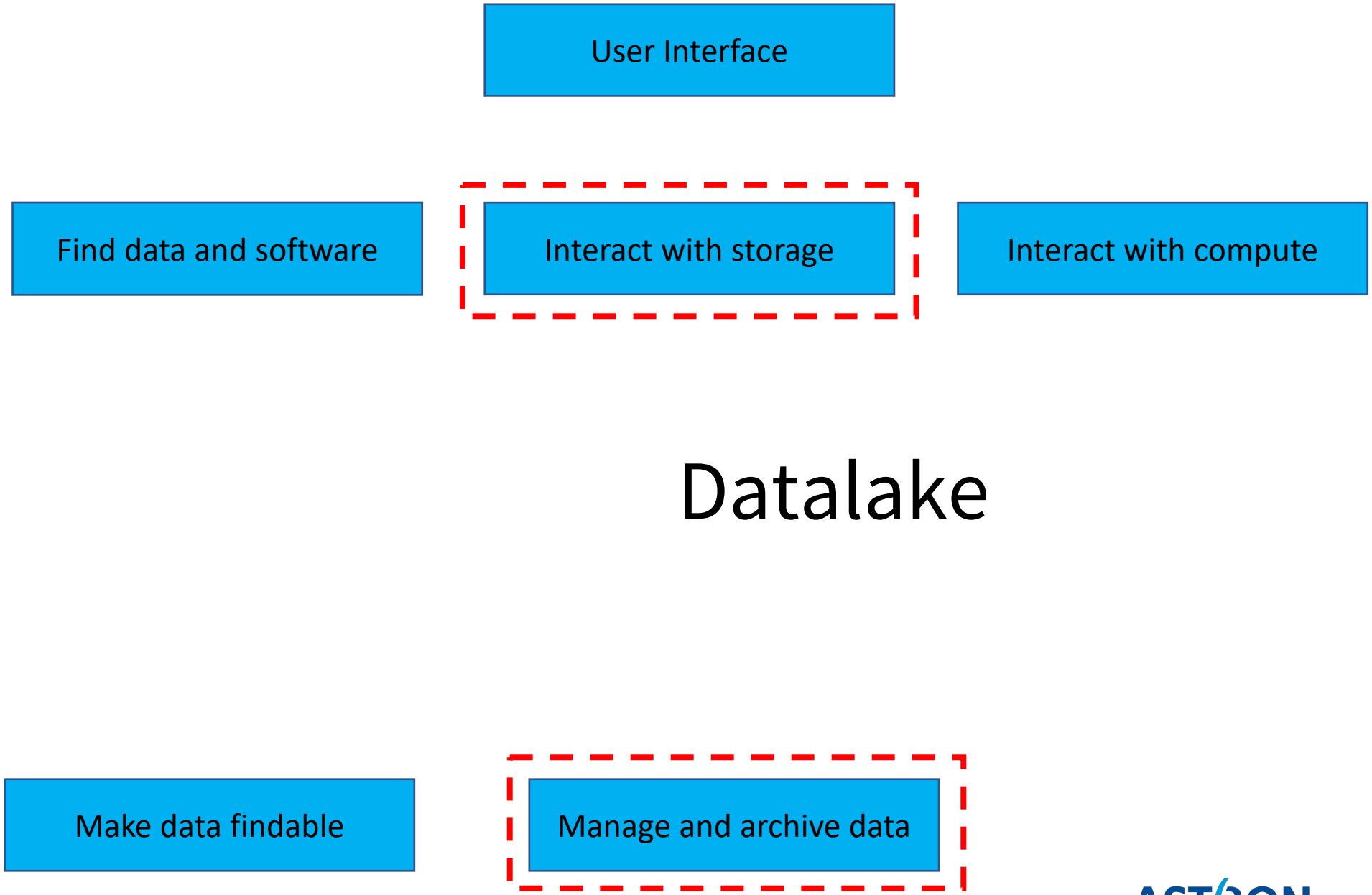
LOFAR LTA data

The data from all LOFAR cycle, commissioning and DDT projects since 2013 are stored in the archive. The interferometric data products that can be found include raw, pre-processed data in the measurement set (MS) format, and the products from the calibration, imaging and long baseline pipelines. In the case of beamformed observations, raw data are available in HDF5 format as well as higher-level data products including de-dispersed time series,

Zooniverse

Zooniverse Classification Database

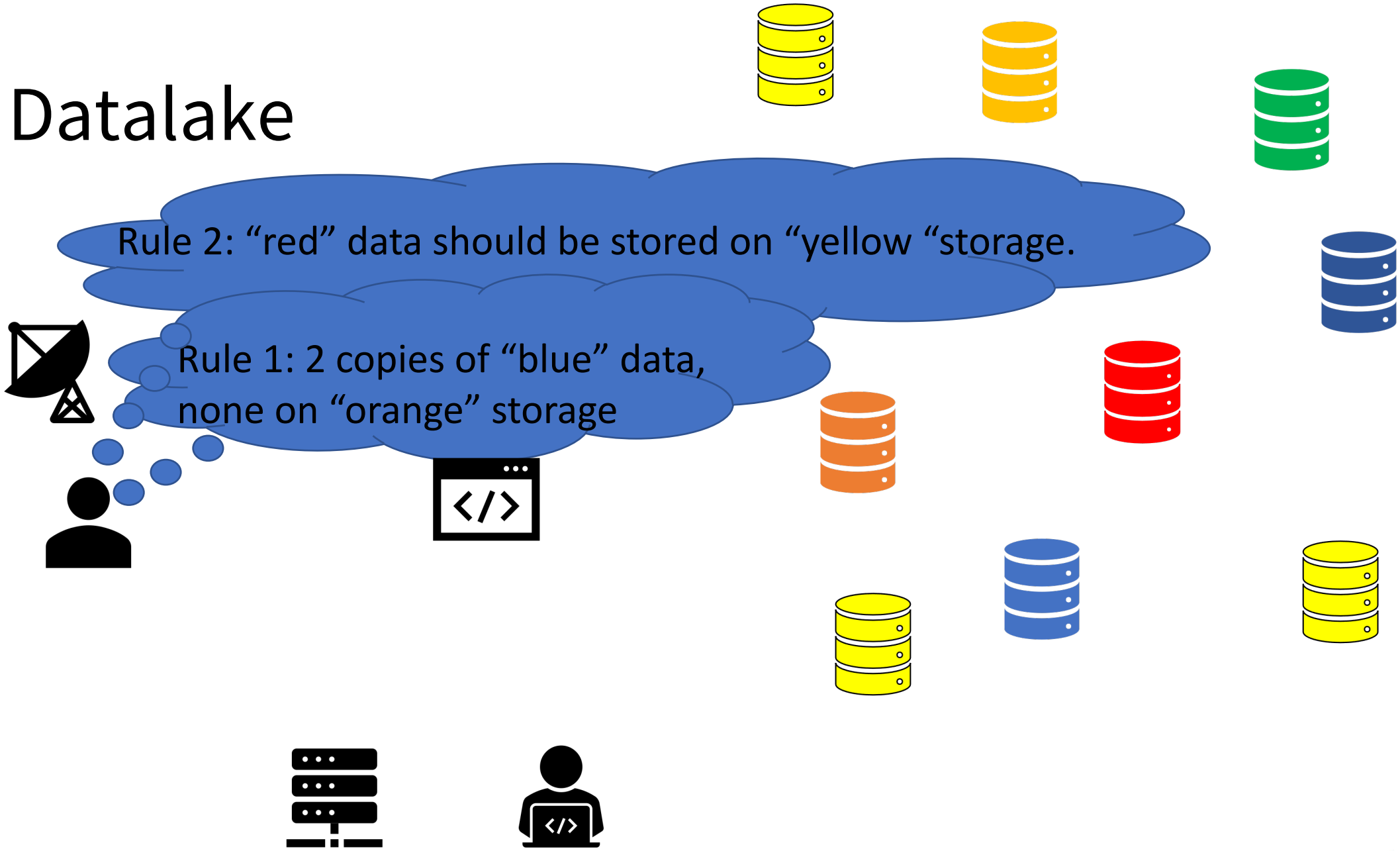
The Zooniverse is the world's largest and most popular platform for people-powered research. This research is made possible by volunteers — more than a million people around the world who come together to assist professional researchers. Our goal is to enable research that would not be possible, or practical, otherwise. Zooniverse research results in new discoveries, datasets useful to the wider research community, and



ASTRON
Netherlands Institute for Radio Astronomy

ESCAPE
European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

Datalake



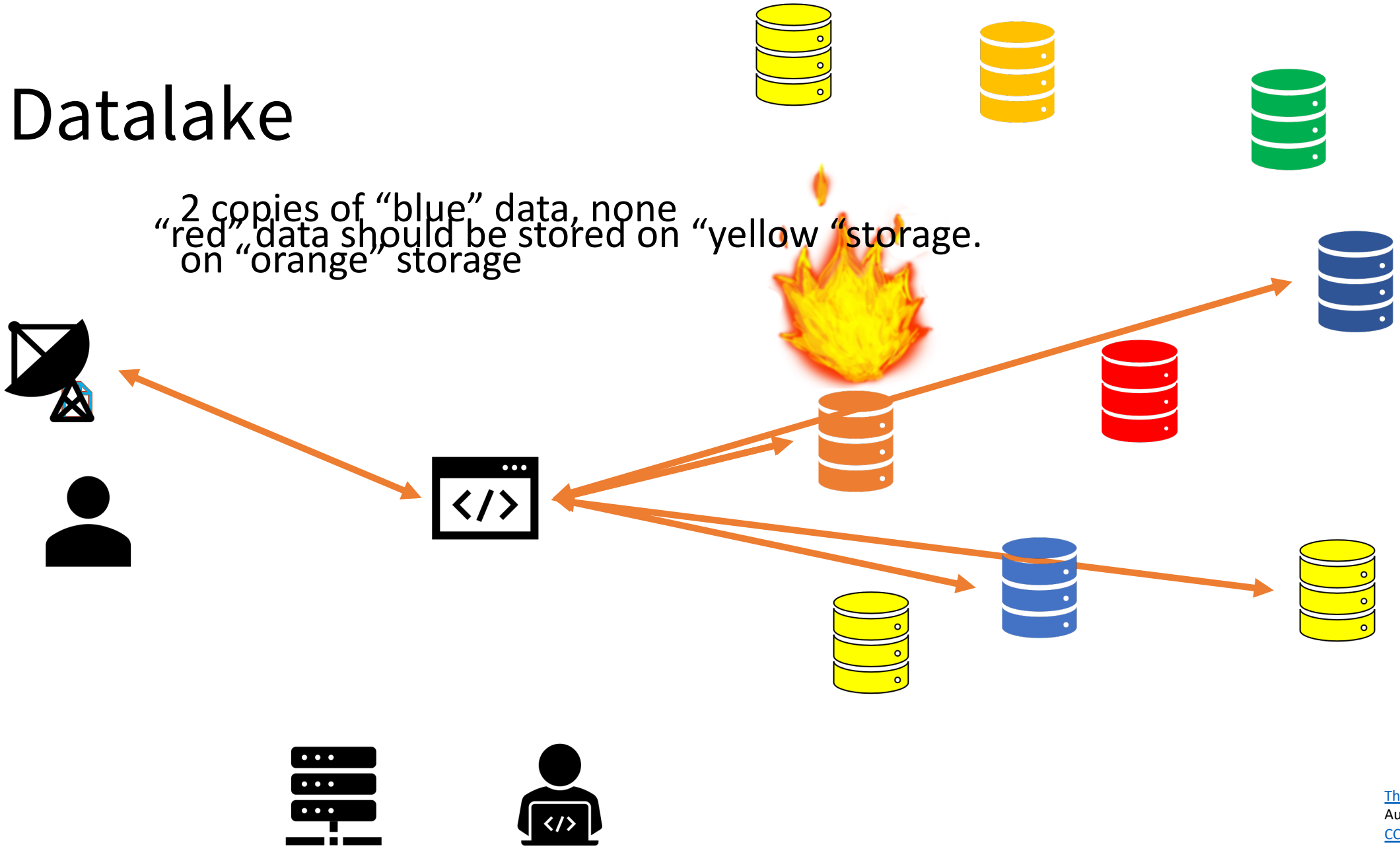
DISCALIMER the author suffers from some colour blindness, if colours do not make sense it’s probably me, not you.

ASTRON Netherlands Institute for Radio Astronomy

ESCAPE European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

Datalake

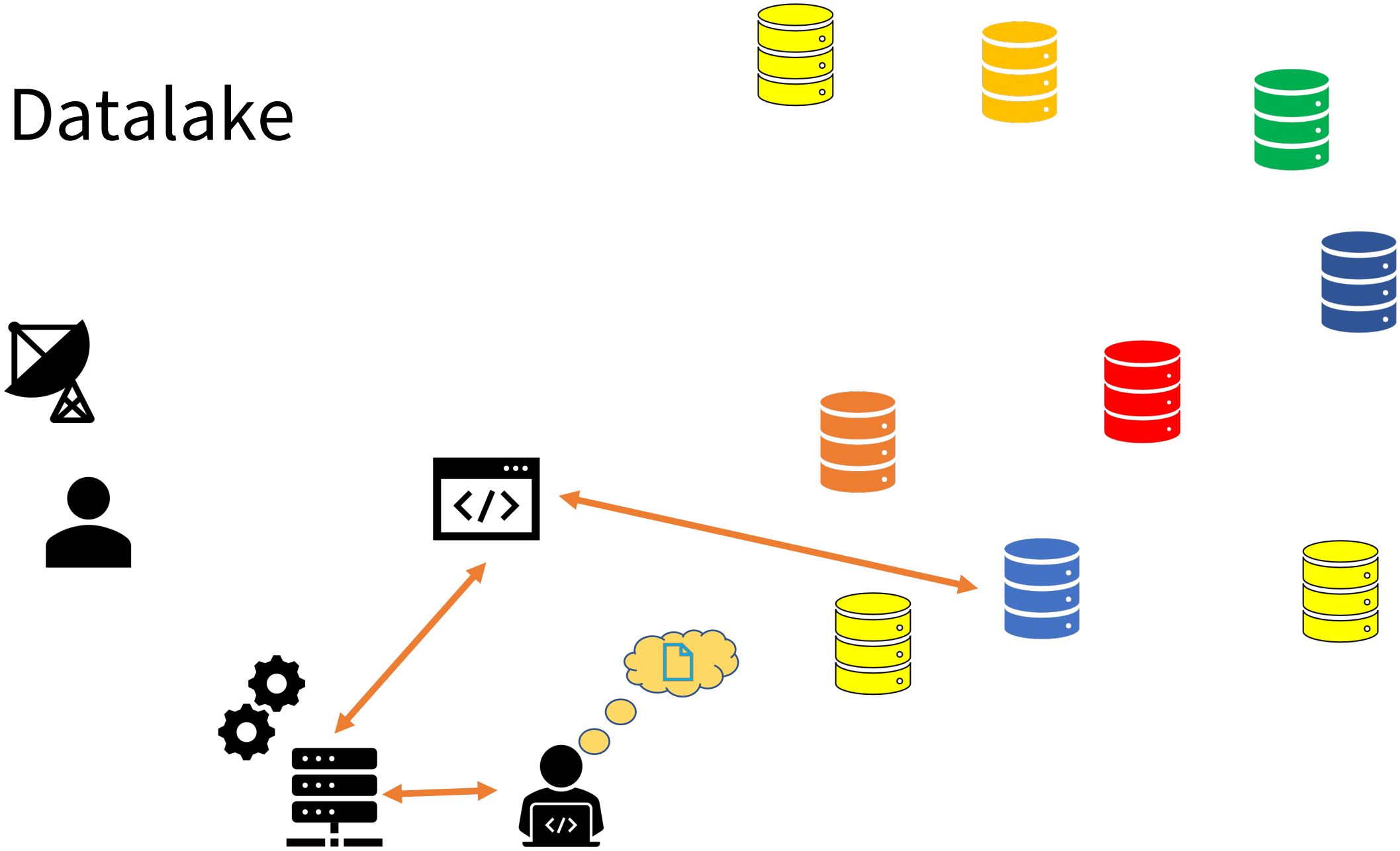
2 copies of “blue” data, none “red” data should be stored on “yellow “storage.
on “orange” storage



This Photo by Unknown Author is licensed under CC BY-SA

DISCALIMER the author suffers from some colour blindness, if colours do not make sense it’s probably me, not you.

Datalake



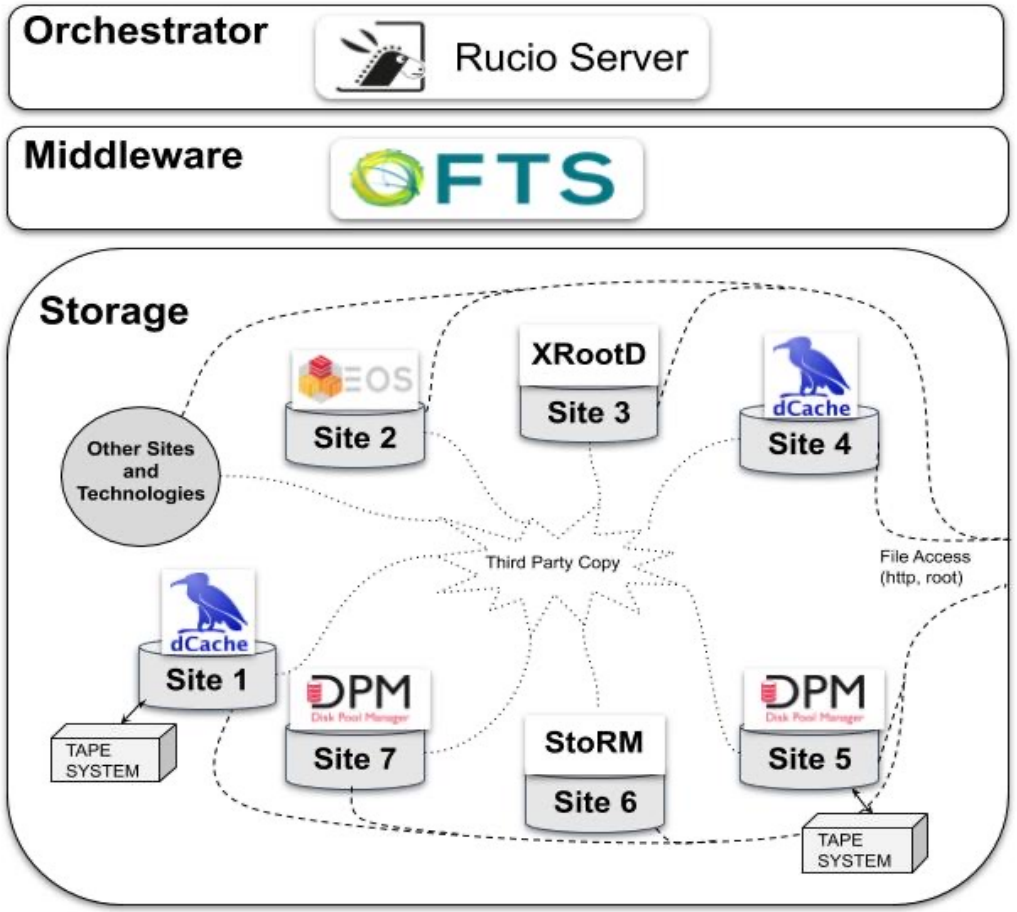
DISCALIMER the author suffers from some colour blindness, if colours do not make sense it's probably me, not you.

ASTRON Netherlands Institute for Radio Astronomy

ESCAPE European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

Datalake

- Investigating the tool chain as used by CERN experiments to manage their data
- Rucio:
  - Rule-based orchecstration
  - CLI, but also python libs, REST, webUI
- FTS:
  - Transfer management



ESCAPE DIOS | Data Infrastructure for Open Science

ASTRON Netherlands Institute for Radio Astronomy

ESCAPE European Science Cluster of Astronomy & Particle physics ESFRI research Infrastructures

Conclusions

- User access and data management are corner stones of the SDC
- The Science Analysis Platform eases the access to data by our users
- The datalake automates the data management from our facilities