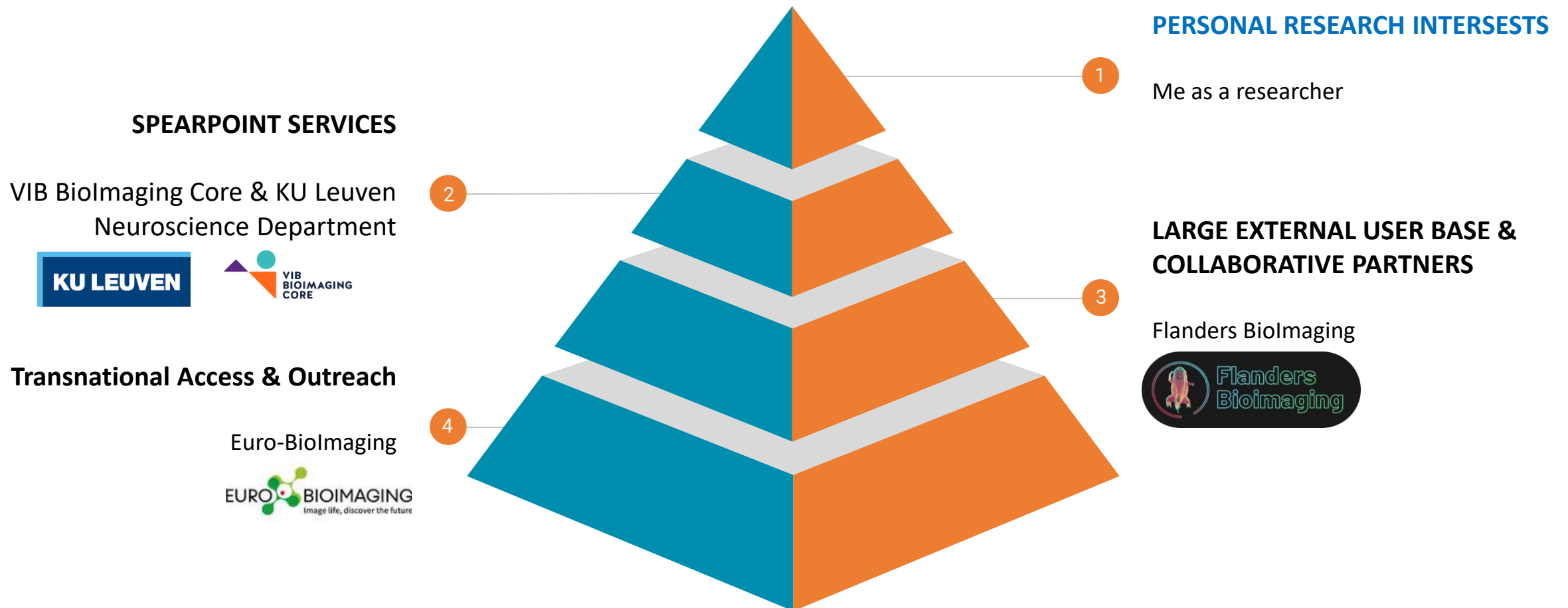


Managing Research Data at Research Infrastructures – The Flanders BioImaging & ManGO connection

Sebastian Munck
VIB Bio Imaging Core
VIB Center for Brain & Disease, Research
KU Leuven, Department of Neurosciences

The 'infrastructure'

Me, the Institute, Flanders BioImaging LIAISE & beyond



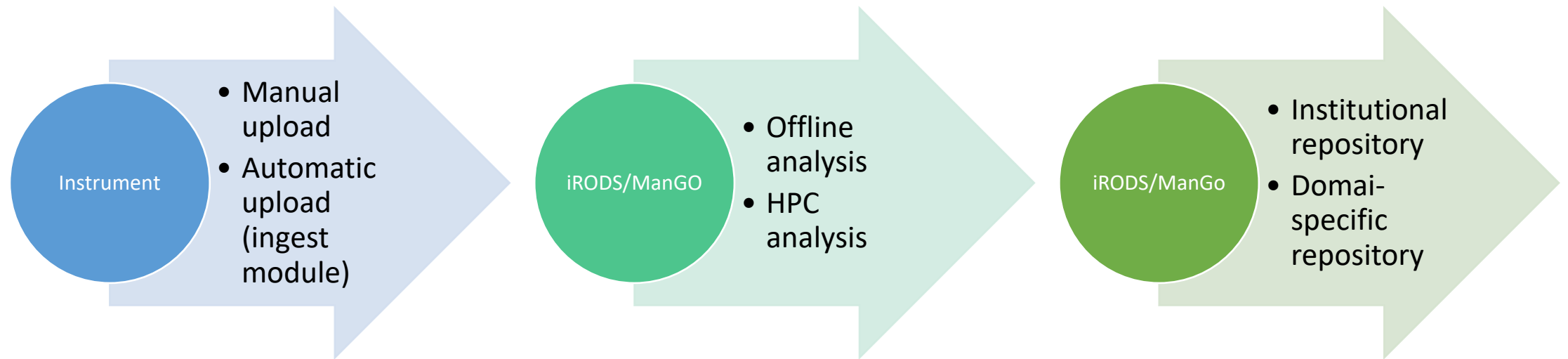
The data problematic

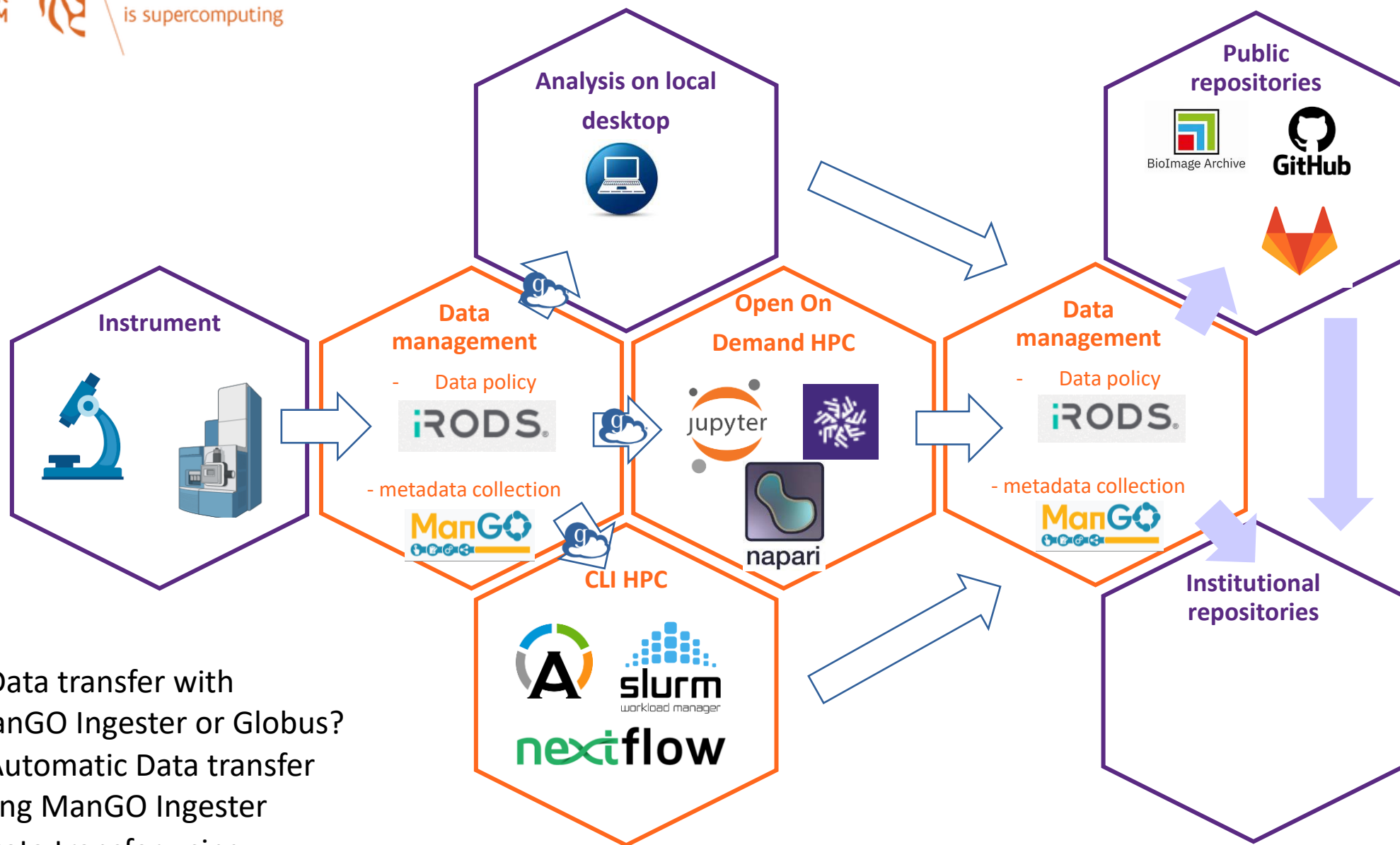
Everyone has big data

- ⇒ Standard detector sCMOS camera 2304×2304 pixels 32-bit/ms
- ⇒ As TIFF = 20,2 MB (21.233.852 bytes)
- ⇒ 20GB/s
- ⇒ 1.2 TB/min
- ⇒ 72TB/h
- ⇒ 1,7PB/day

- Not used like this
- Not all is information
- Most is background
- Different analysis challenges

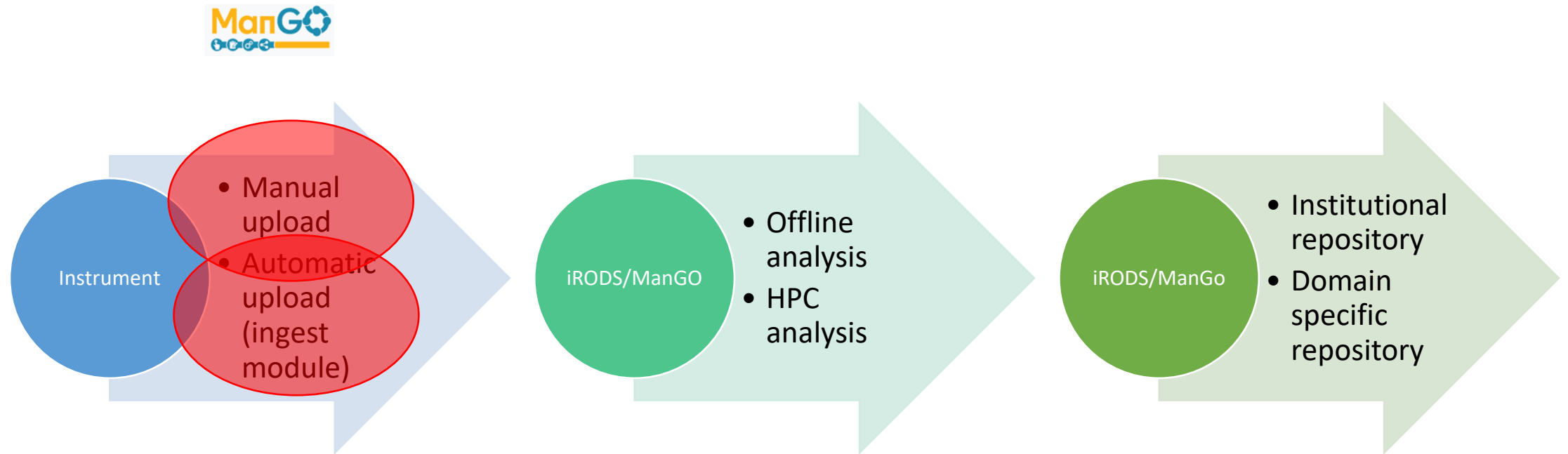
The concept





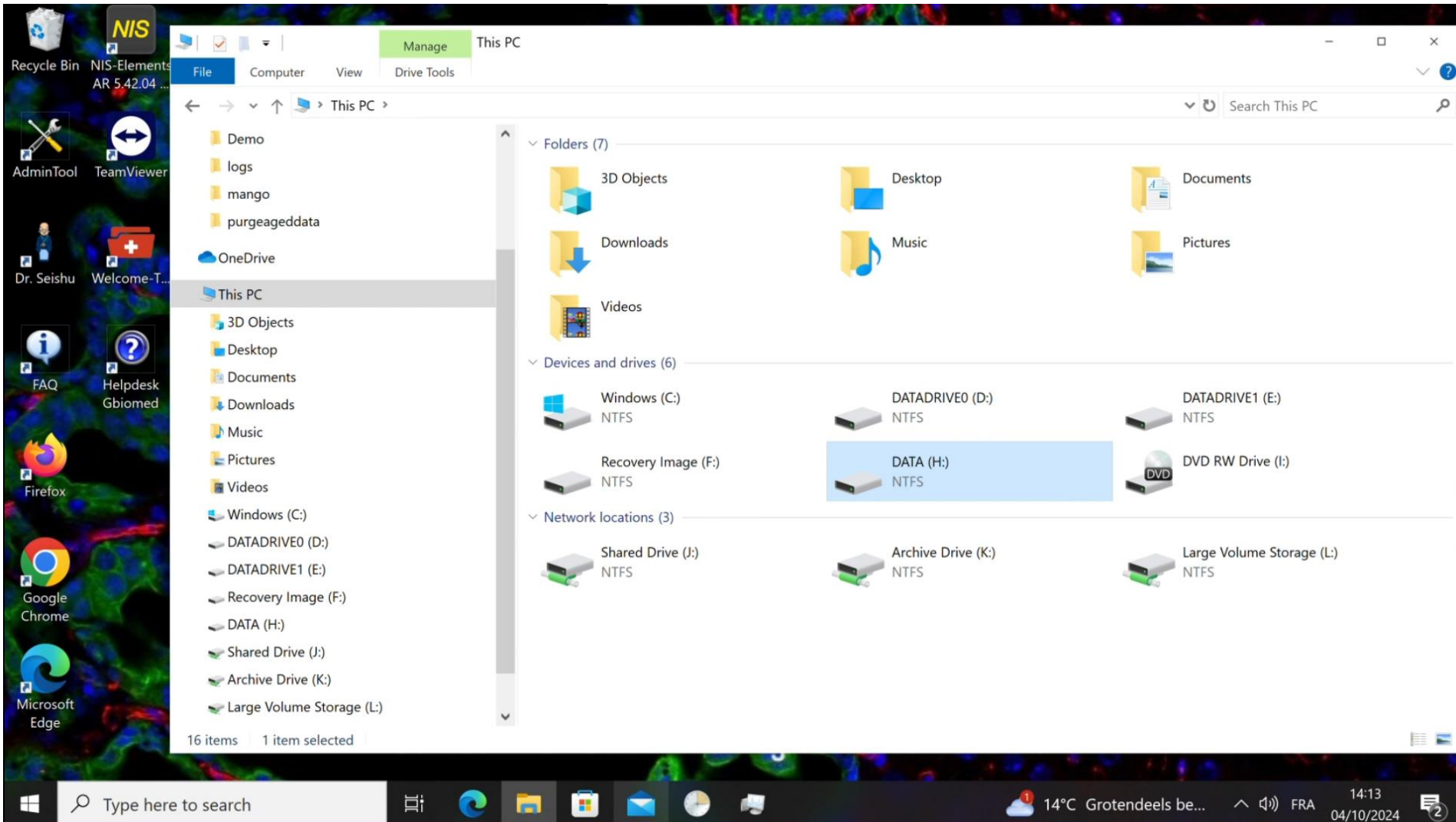
- = Data transfer with ManGO Ingester or Globus?
- = Automatic Data transfer using ManGO Ingester
- = Data transfer using Globus

Transferring data to iRODS



ManGO job creator - making automatic upload a reality

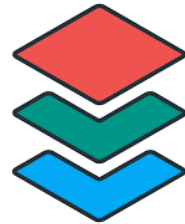
proof of concept



- Work in progress
- Part of the bigger workflow
- Means new training for users
- Prevents the clogging of the local hard drives
- Managed process

Understanding the file formats is important to make it work for all our machines

- Absence of a standard file format makes life hard
- All suppliers have different file formats
- File formats of individual suppliers are complex
- Bioformats a community standalone Java library for reading and writing life sciences image file formats has limits



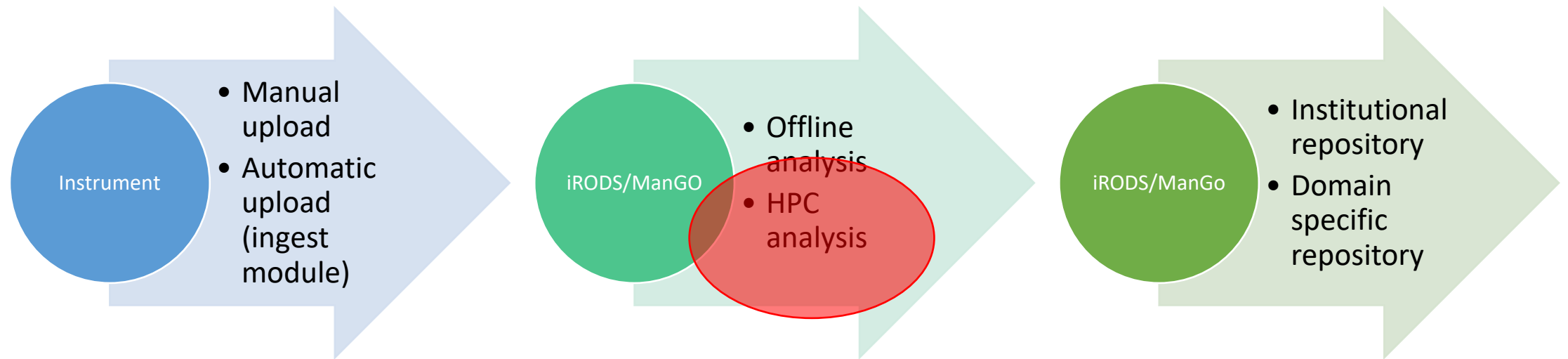
- Info on file reader
- .nd2-reader (Nikon file format)
- Work in progres

```
current_dir=os.getcwd
df = pd.DataFrame(columns=["filename", "microscope", "camera", "objective", "channels", "software"])
for i in os.listdir(current_dir()):
    if i.endswith('.nd2'):
        with ND2FileExt(i) as f:
            df = df._append({'filename': i,
                             'microscope': f.mic_name,
                             'camera': f.cam_name,
                             'objective': f.obj_name,
                             'channels': f.chan_name,
                             'software': f.sw_name
                             }, ignore_index=True)
df
```

✓ 0.8s Python

	filename	microscope	camera	objective	channels	software
0	ax11.nd2	Nikon Ti2	Nikon_Confocal_Ax	Plan Apo Lambda S 40XC Sil	TRITC (Ex:561.0nm, Em:571.0nm)	NIS-Elements AR 6.02.03 (Build 1993)
1	cicero6.nd2	Nikon Ti2	Prime BSI Express A23H726052	Plan Apo Lambda S 25XC Sil	TRITC (Ex:561.0nm, Em:594.0nm)	NIS-Elements AR 6.02.03 (Build 1993)
2	L2135_saponin_Reconstructed.nd2	Nikon Ti2	Hamamatsu C11440-22C SN:303144	SR Apo TIRF AC 100xH	3D-SIM_640 (Ex:Nonenm, Em:701.5nm), 3D-SIM_561...	NIS-Elements AR 5.30.07 (Build 1569)
3	NikonTi2_AX_40xSil.nd2	Nikon Ti2	AX	Plan Apo Lambda S 40XC Sil	TRITC (Ex:561.0nm, Em:571.0nm)	NIS-Elements AR 5.42.06 (Build 1821)
4	no2B SODG93A 5M FXYD6 20xZstack new set001.nd2	Nikon Ti	Nikon C2plus	Plan Apo VC 20x DIC N2	Alexa Fluor 647 dye-labeled oligonucleotide/H2...	NIS-Elements AR 5.21.03 (Build 1489)

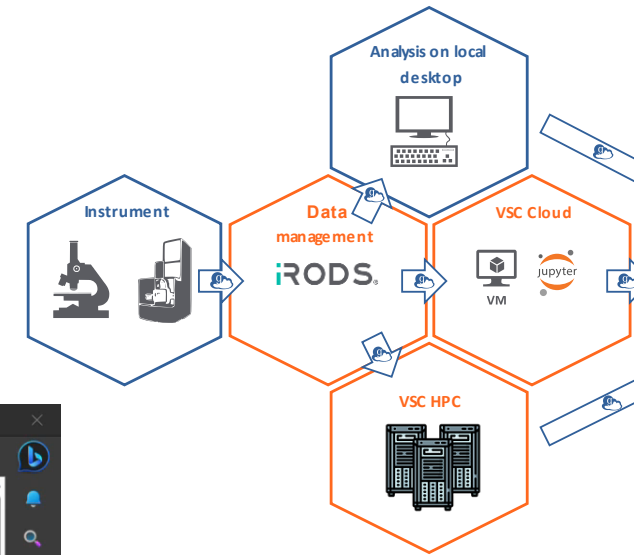
Now that we uploaded the data – Connecting with analysis



Analysis using Open-OnDemand



Till Korten
& Tatiana



KU LEUVEN

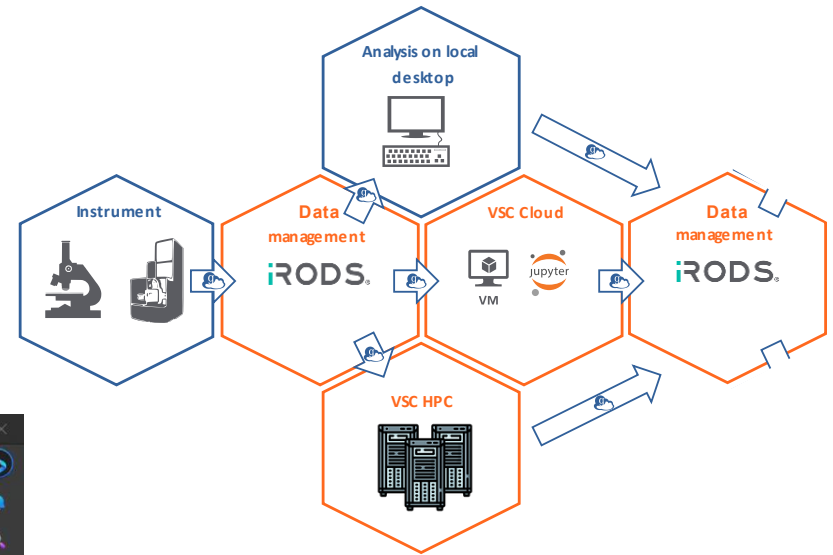
OnDemand provides an integrated, single access point for all of your HPC resources.

Pinned Apps A featured subset of all available apps

- Active Jobs System Installed App
- Home Directory System Installed App
- Job Composer System Installed App
- Login Server Shell Access System Installed App
- code-server System Installed App
- Interactive Shell System Installed App
- Jupyter Lab System Installed App
- RAPIDS Nvidia Rapids System Installed App
- RStudio Server System Installed App
- Tensorboard System Installed App

- Running a Jupyter notebook on the VSC to analyze the uploaded data

From the analysis to ManGO



```
[ ]: summary = cell_measurements.groupby('method')[['number_of_cells', 'area', 'intensity_mean']].mean()
summary
[ ]: summary.to_csv("summary.csv")
[ ]: cell_measurements.to_csv("cell_measurements.csv")

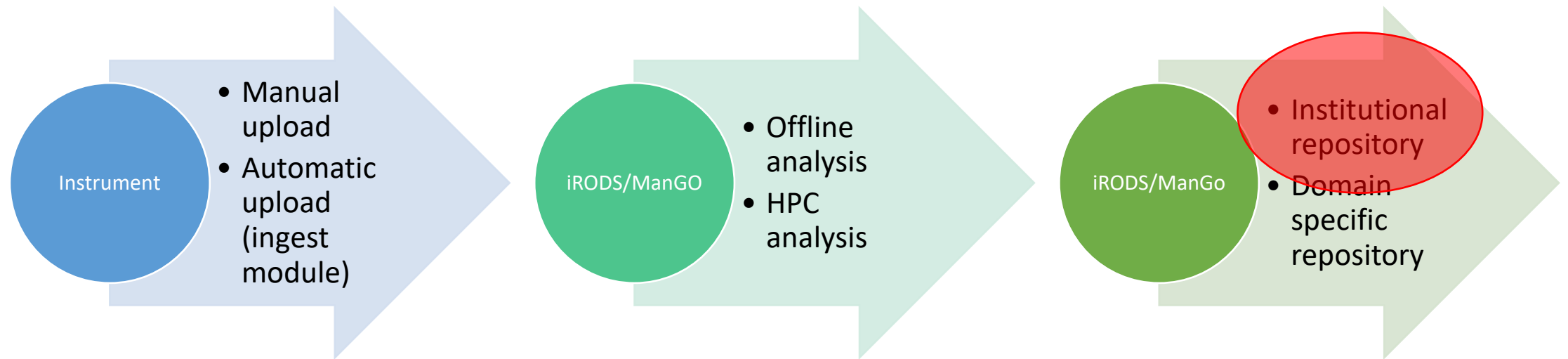
5. Export metadata to ManGO using iRODS
Connection to ManGO using irods-pythonclient

[1]: config={
    "irods_host": "gbiomed-fbi.irods.icts.kuleuven.be",
    "irods_port": 1247,
    "irods_zone_name": "gbiomed_fbi",
    "irods_authentication_scheme": "PAM",
    "irods_encryption_algorithm": "AES-256-CBC",
    "irods_encryption_salt_size": 8,
    "irods_encryption_key_size": 32,
    "irods_encryption_num_hash_rounds": 8,
    "irods_user_name": "u0152370",
    "irods_ssl_ca_certificate_file": "",
    "irods_ssl_verify_server": "cert",
    "irods_client_server_negotiation": "request_server_negotiation",
    "irods_client_server_policy": "CS_NEG_REQUIRE",
    "irods_default_resource": "default",
    "irods_cwd": "/gbiomed_fbi/home",
    "irods_authentication_uid": 1000
}

[2]: password = "64da6a06.12zb1D11lopm1QZYH1Q6WRCnp2YQuq3sxYCH"
from irods.session import iRODSSession
from irods.password_obfuscation import encode
from irods.meta import iRODSMeta
import os
def iinit(config, password):
    def put(file, contents):
        os.makedirs(os.path.dirname(file), exist_ok=True)
        with open(file, "w") as f:
            f.write(contents)
    env_file = os.getenv('IRODS_ENVIRONMENT_FILE', os.path.expanduser('~/.irods/irods_environment_python.json'))
    put(env_file, config)
```

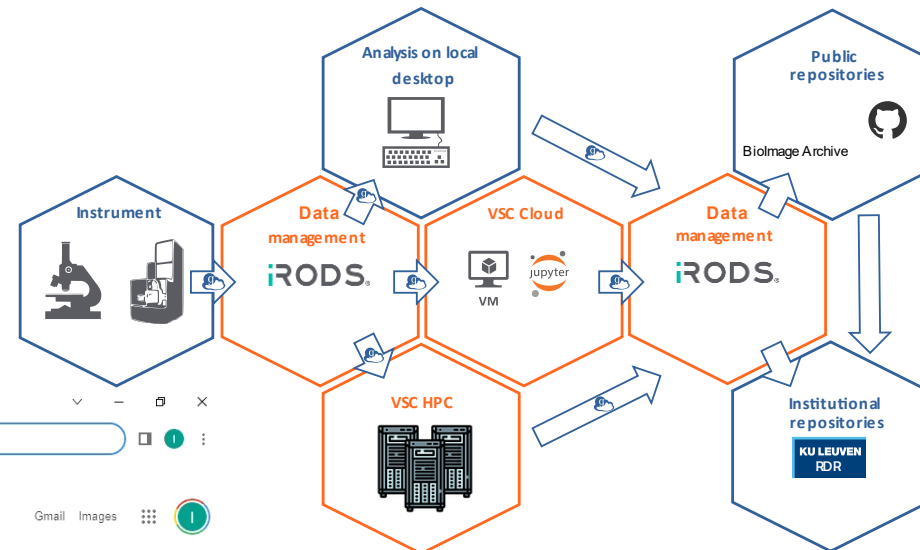
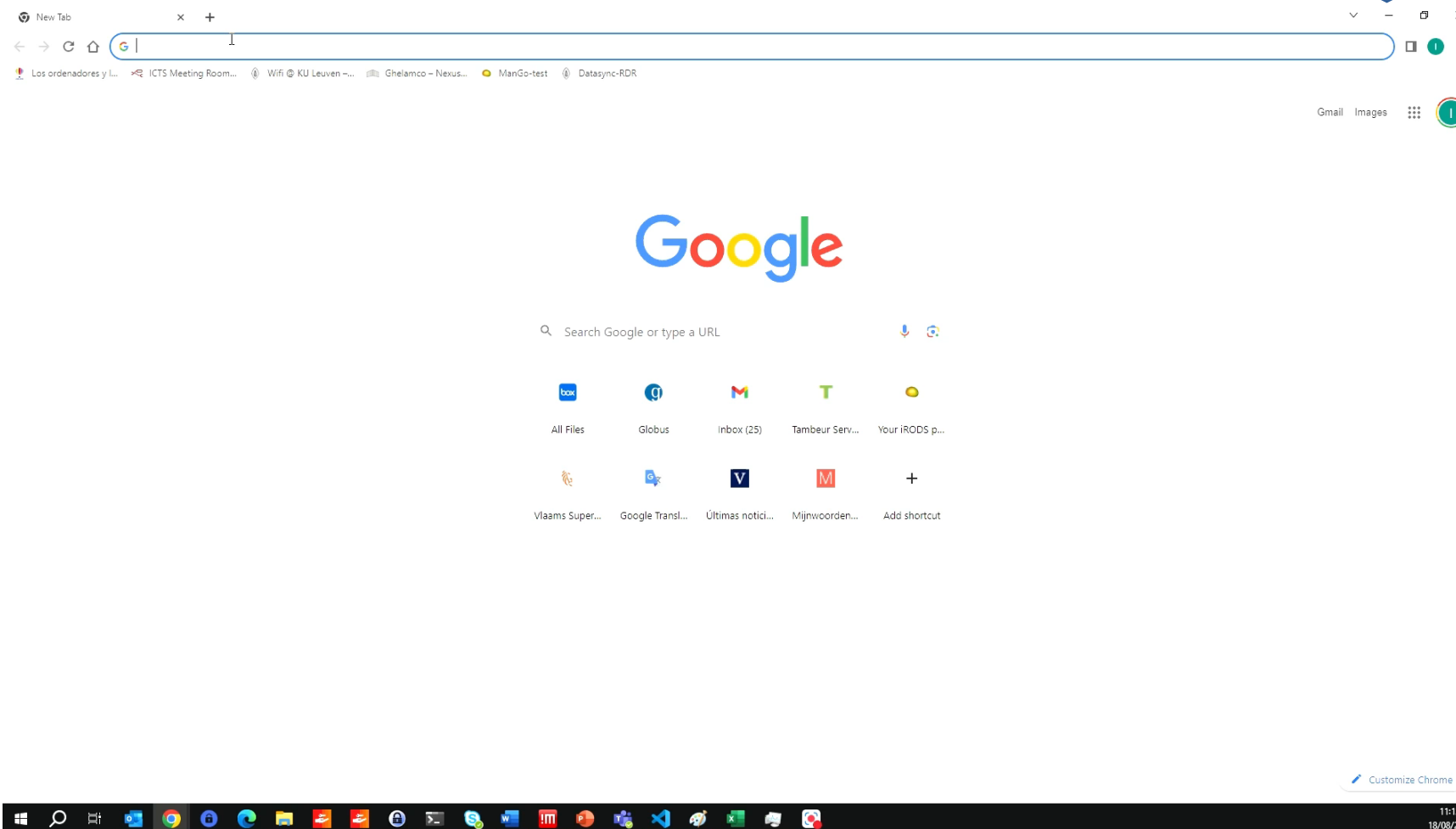
- Connecting to ManGO using a python irods-client.
- Checking that the file to which we add metadata is present
- Create a new folder, “analysis,” in ManGO and export the csv file together with some metadata.
- On the ManGO platform checking whether the csv file is present and if it has the right metadata.

Publicizing the data



Publish to an institutional repository

proof of concept



- Create a dataset on RDR (the KU Leuven data repository)
- Use the RDR Integration tool to directly select and upload files to RDR

Local repository to publish research data

KU Leuven RDR

RDR (pronounced "Radar") is **KU Leuven's institutional research data repository for the publication of research data**. A Dataverse.org based platform to upload, describe, and share your research data to make your data more FAIR.

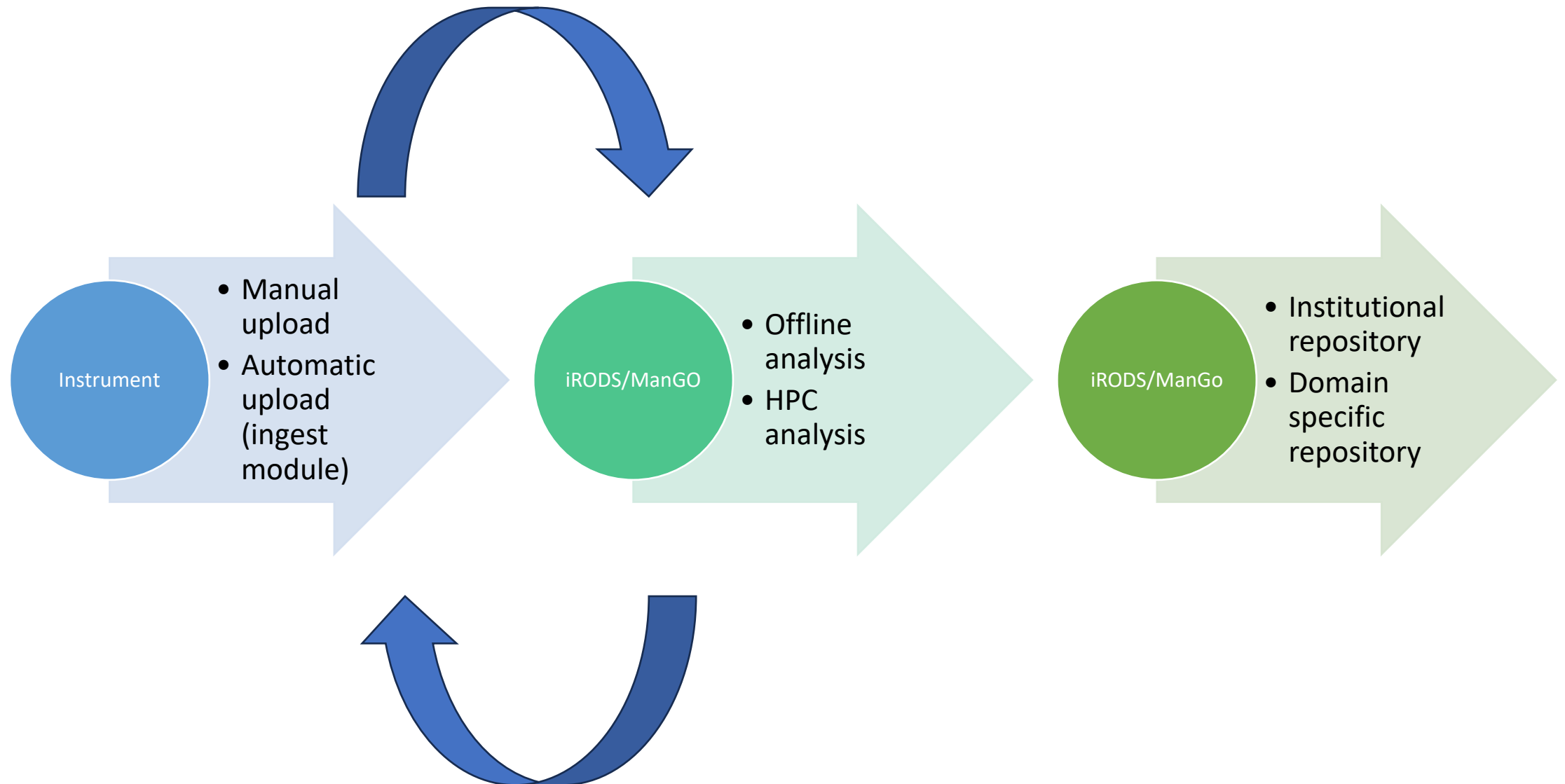
Make your data citeable: A published dataset in KU Leuven RDR gets its own DOI, is registered in Lirias, and appears on your who-is-who publication list.

[User Manual](#)[Integration Dashboard](#)

- Future direct upload from ManGo to other repositories including the Biolmage Archive

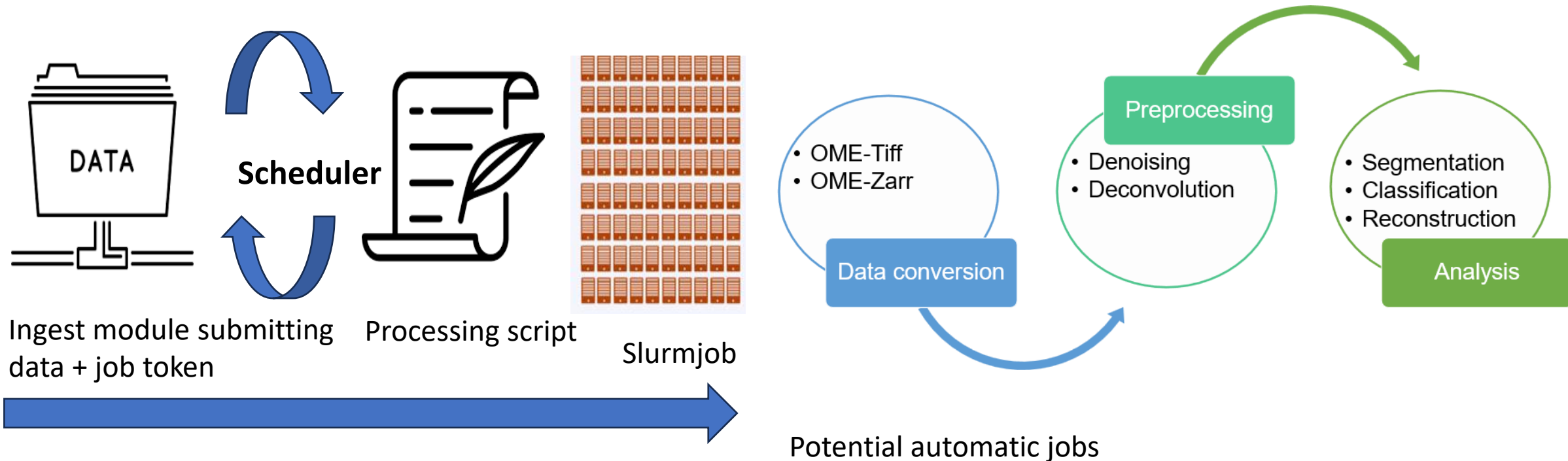
Individual analysis (like in these jupyter notebooks) is great – but how about automating part of the processing in a ‘blind’ fashion?

Automating more steps in the workflow



SCROn-jobs: Work in progress

- A **cron job** is a Linux command used for scheduling tasks to be executed sometime in the future. This is normally used to schedule a job that is executed periodically – for example, to send out a notice every morning. Scron jobs are targeted at Slurm clusters



Automatic ingestion



Format conversion metadata schema:

- To OME-TIFF: pending | done
- To OME-ZARR: pending | done

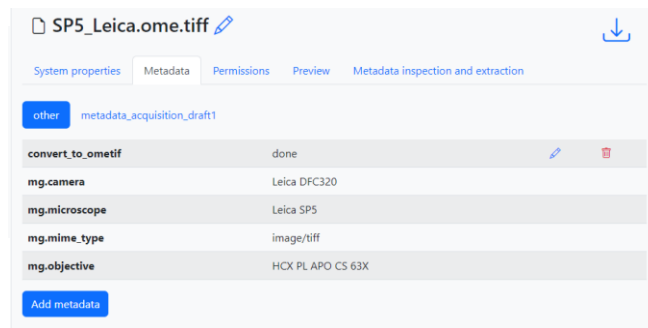
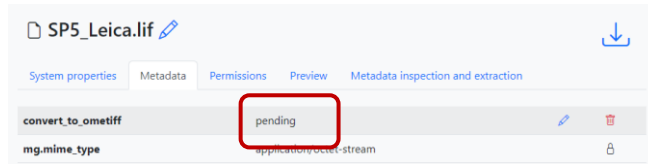


.lif files

The job will:

- Search file based on metadata
- Convert to OME-TIFF
- Upload to ManGO
- Add relevant metadata

Scronjob -daily

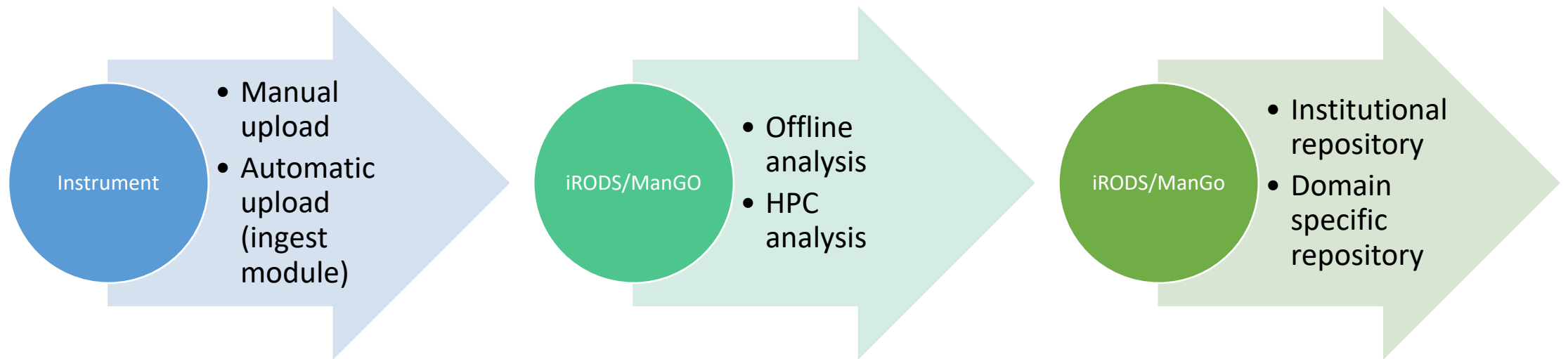


.tiff files



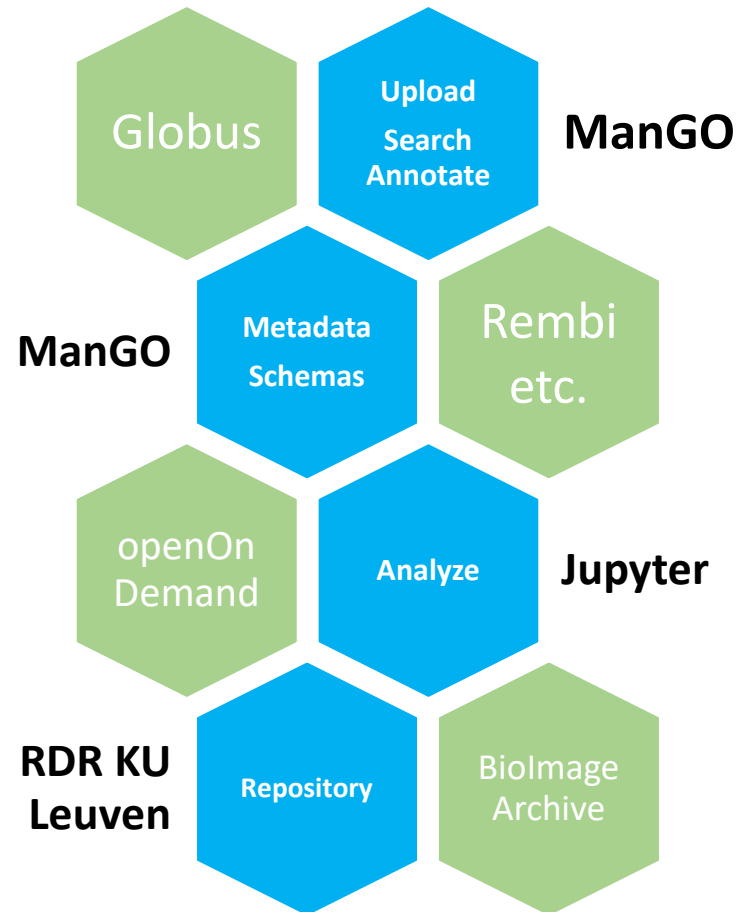
BatchConvert

A managed workflow



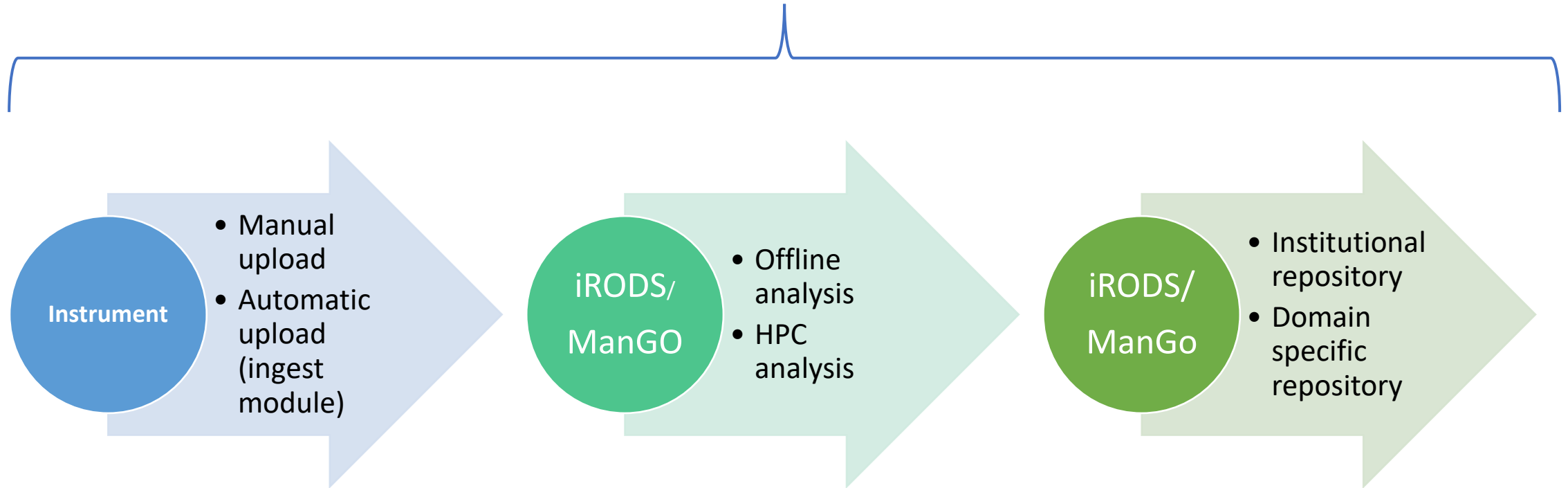
Quick summary: The data journey through connected tools

FAIR by design



- Transfer
 - RDM
 - Analysis
 - Deposition
- Open source
 - Complete workflow from acquisition to publishing
 - Flexible
 - Content agnostic
 - Using robust pieces that click together
 - ManGO is central

Integrating the workflow in the bigger environment



It is a complex environment

My metadata



My Electronic lab-notebook

Initiatives, stakeholders, and Funders



Meta data standards

This block contains a collage of scientific articles and guidelines related to microscopy metadata standards:

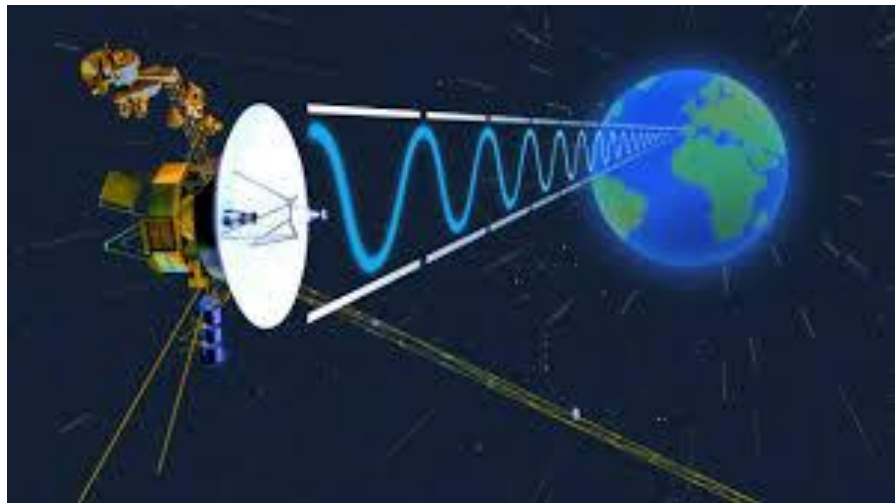
- REMBI: Recommended Metadata for Biological Images—enabling reuse of microscopy data in biology** (Comment Focus)
- Community-developed checklists for publishing images and image analyses** (Nature Methods)
- Microscopy-BIDS: An Extension to the Brain Imaging Data Structure for Microscopy Data** (Journal of Microscopy)
- Standard metadata for 3D microscopy** (Journal of Microscopy)
- Microscopy-Meta App: an interactive tool for collecting microscopy metadata based on community specifications** (Comment Focus)
- FAIR High Content Screening in Bioimaging** (Open Access)
- Towards community-driven metadata standards for light microscopy: tiered specifications extending the OME model** (Comment Focus)
- MITI minimum information guidelines for highly multiplexed tissue images** (Comment Focus)

So there are many stakeholders, but who is in charge?

...you?

How do we combine the *scattered* information and correct errors?

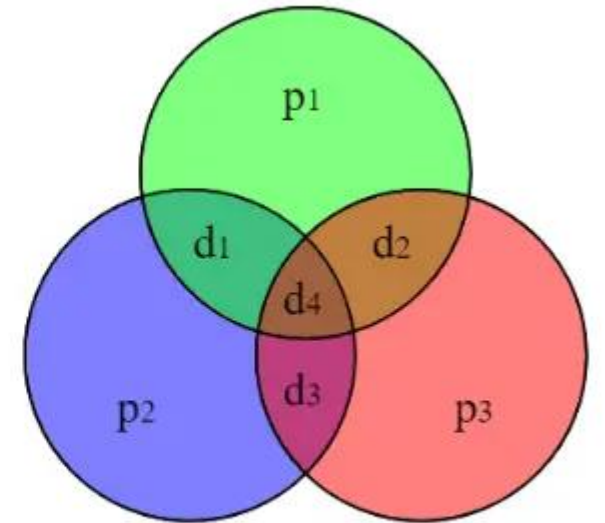
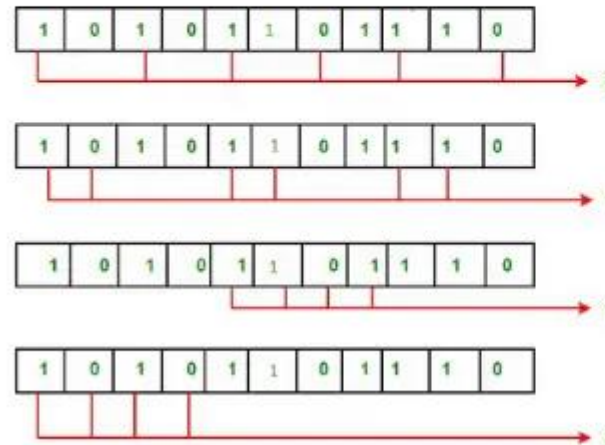
Maybe we can use the redundancy to our advantage



Communication is also difficult in other disciplines

Data can be transported over noisy channels (Shannon), and it can be corrected (Hamming)

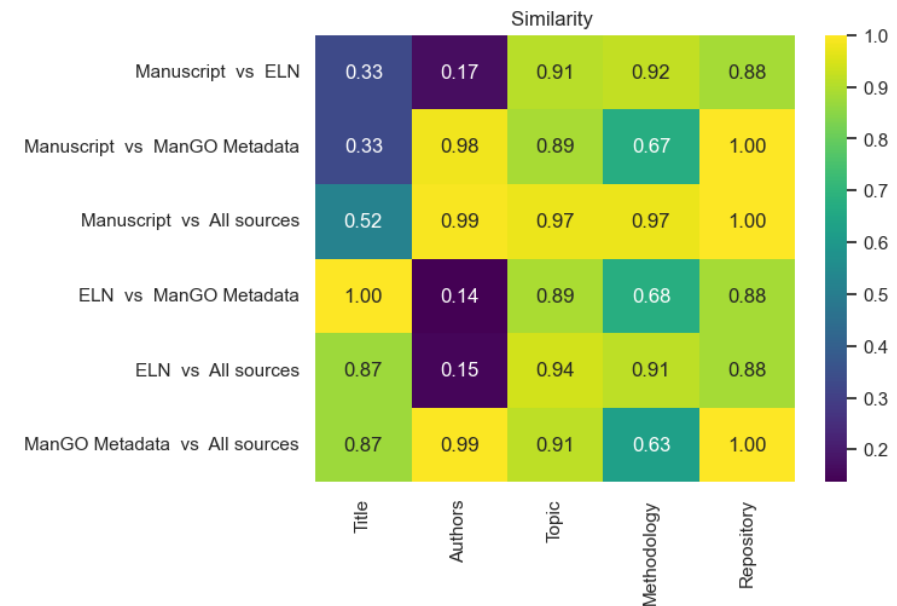
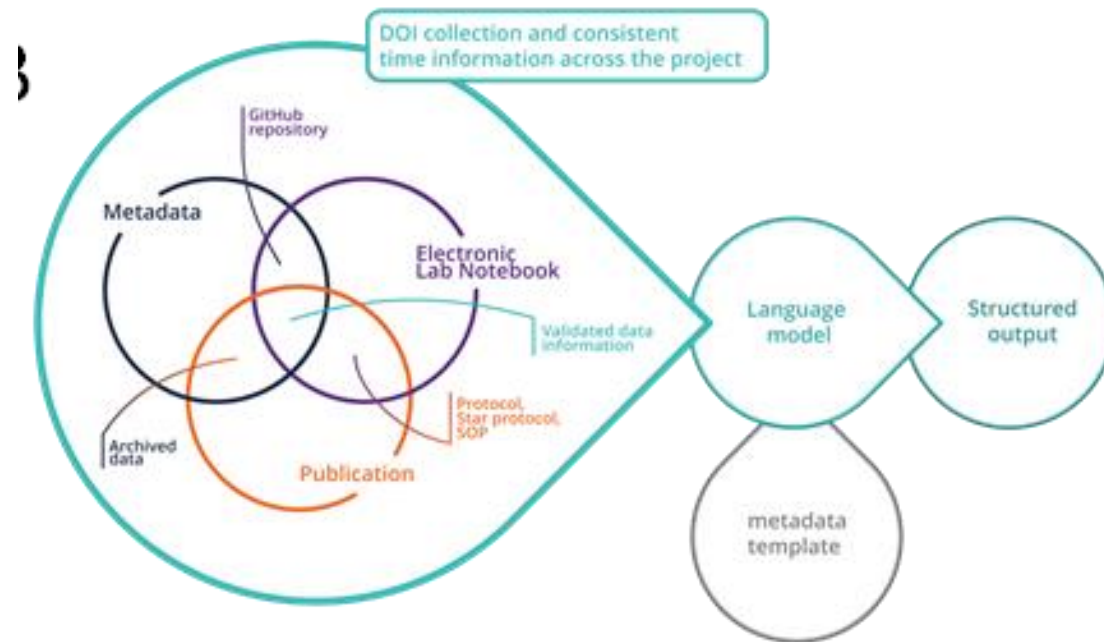
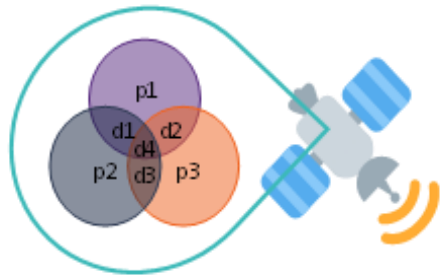
What is Hamming Code?



Electrical 4 U

The scattered information can be digested by ChatGPT and the differences visualized

- Learning from deep space communication for reproducible BioImaging and data analysis
- Using information from lab notebooks, manuscripts, and meta-data servers can be seen similarly as redundant overlapping information



Summary

- A lot of what I showed is work in progress
- Managing Research Data at Research Infrastructures – The Flanders BioImaging & ManGO connection – is complex and multifaceted
- Automation and scale-up is your friend
- The tools available with KU Leuven and the VSC are powerful
- ManGo is central, and only the beginning

FLANDERS BIOIMAGING - LIAISE

Leading Imaging Application
Integrated Service Enablement

find me @

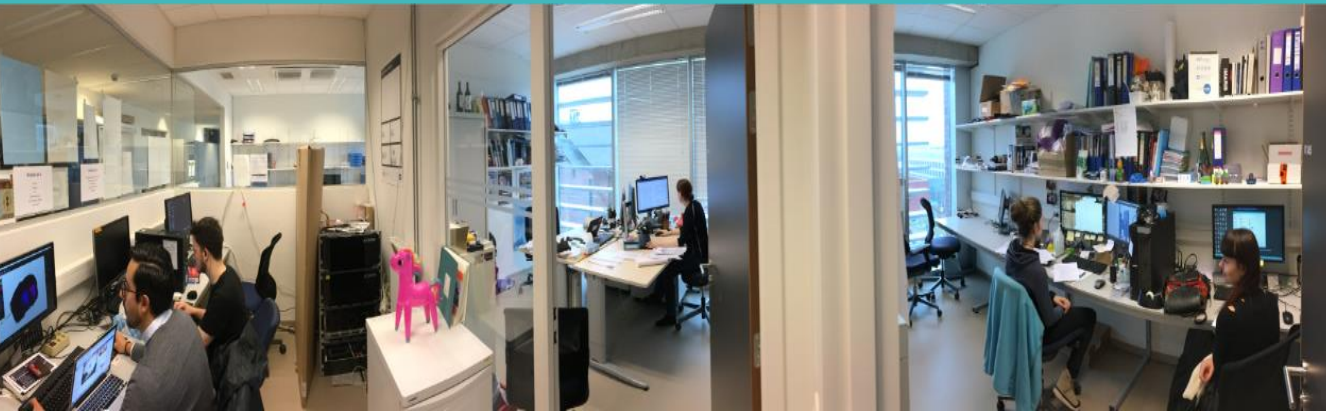
VIB-KU LEUVEN

CENTER FOR BRAIN
& DISEASE RESEARCH

cbd.vib.be



@BioImagingCore; @SebastianMunck



KU LEUVEN



Thanks to:

All labs and users

VIB Bio Imaging Core Facility

Benjamin Pavie
Nicolas Peredo
Natalia Gunko
Pieter Baatsen
Katlijn Vints
Abril Escamilla Ayala
Nikky Corthout
Axelle Kerstens
Hélène Roberge

University of Antwerp

Winnok De Vos
Marlies Verschuuren
Tim Van De Looverbosch

KU Leuven Nuclear medicine & molecular imaging

Koen Van Laere
Chris Cawthorne

VIB Technology Training

Alex Botzki
Tatiana Woller

VSC

Ingrid Barcena-Roig
Jan Ooghe
Mariana Montes

TU Dresden

Till Korten
Robert Haase

