SCIENCE MEETS LIFE

CENTER of EXCELLENCE
supported by Nikon

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

VIB

KATHOLIEKE UNIVERSITEIT LEUVEN

The 'infrastructure'

Me, the Institute, Flanders BioImaging LIAISE & beyond

SPEARPOINT SERVICES

VIB BioImaging Core & KU Leuven Neuroscience Department

Transnational Access & Outreach

Euro-BioImaging

PERSONAL RESEARCH INTERSESTS

Me as a researcher

LARGE EXTERNAL USER BASE & COLLABORATIVE PARTNERS

Flanders BioImaging

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

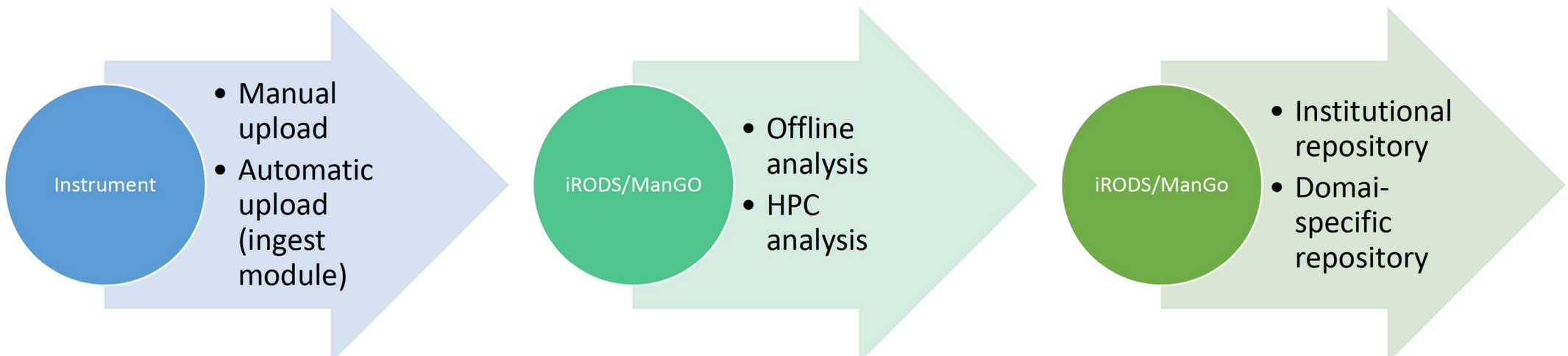
⇒ 1.2 TB/min

⇒ 72TB/h

⇒ 1,7PB/day

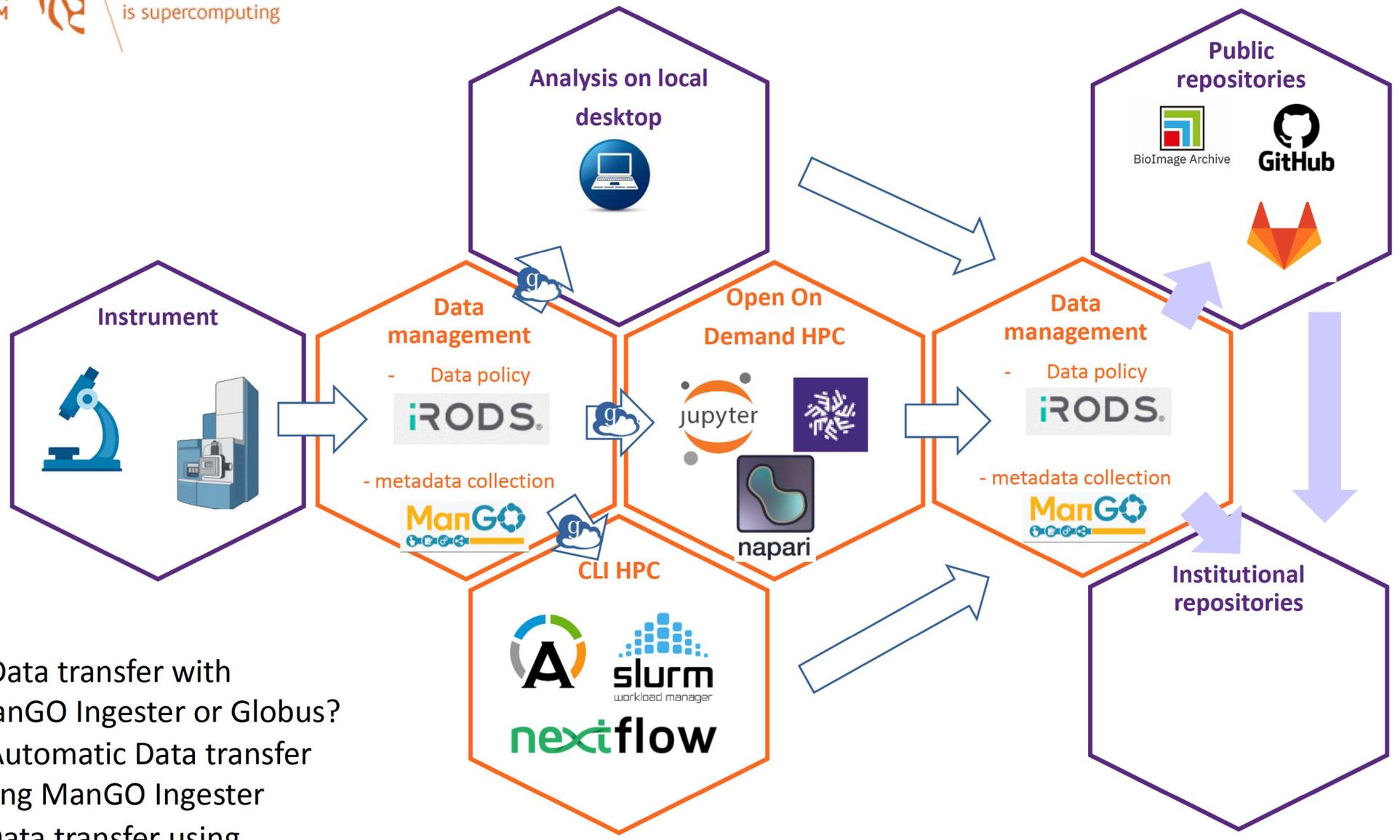
- Not used like this
- Not all is information
- Most is background
- Different analysis challenges

The concept


Instrument
• Manual upload
• Automatic upload (ingest module)
iRODS/ManGO
• Offline analysis
• HPC analysis
iRODS/ManGo
• Institutional repository
• Domai-specific repository

VLAAMS SUPERCOMPUTER CENTRUM

Vlaanderen is supercomputing

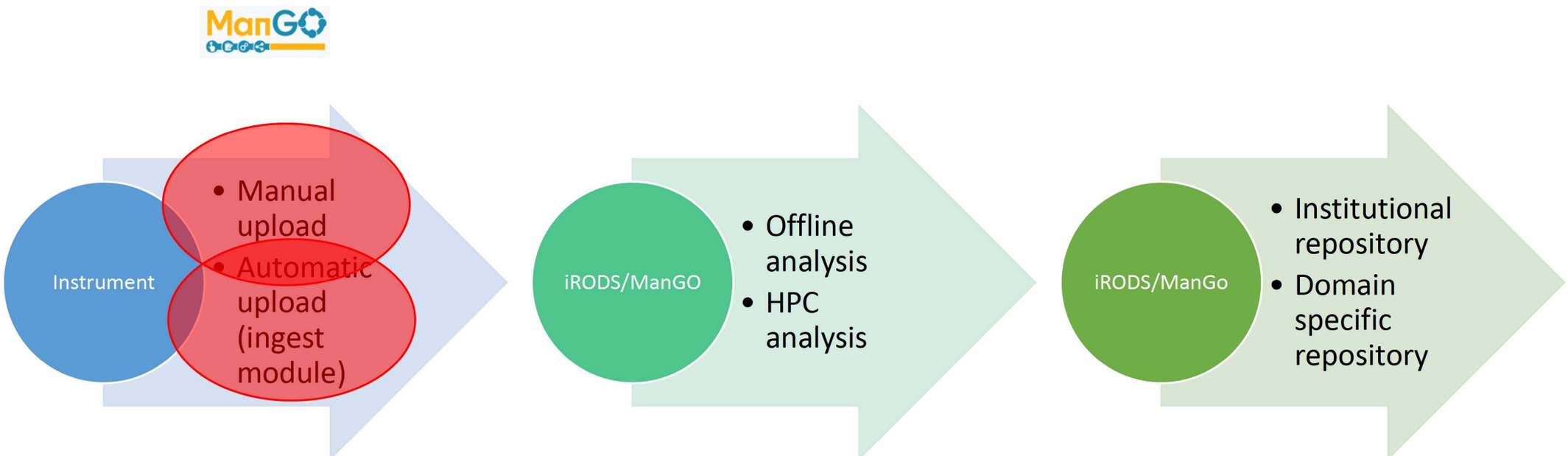


= Data transfer with ManGO Ingester or Globus?

= Automatic Data transfer using ManGO Ingester

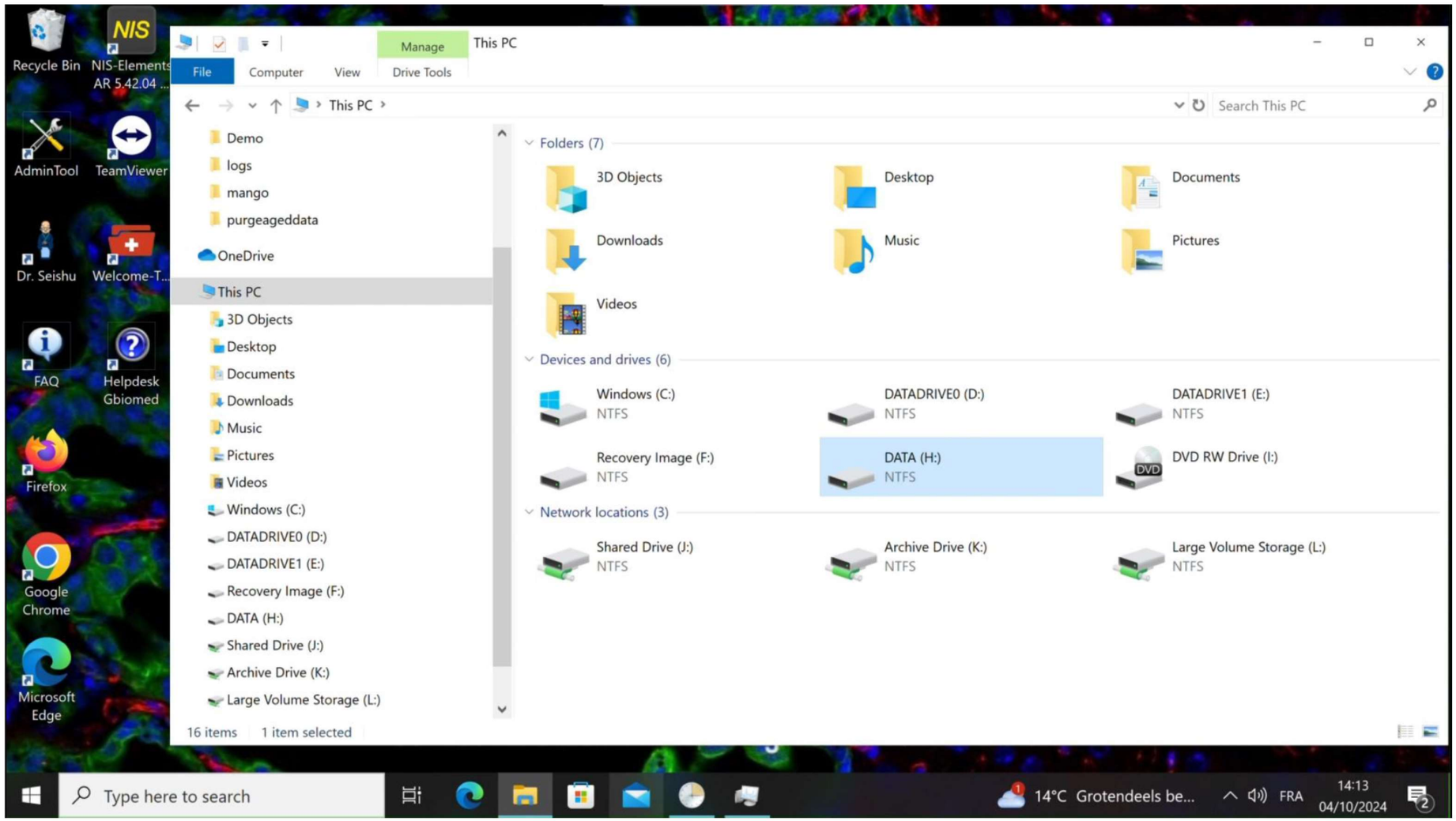
= Data transfer using Globus

Transferring data to iRODS



ManGO job creator - making automatic upload a reality

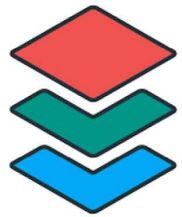
proof of concept



- Work in progress
- Part of the bigger workflow
- Means new training for users
- Prevents the clocking 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



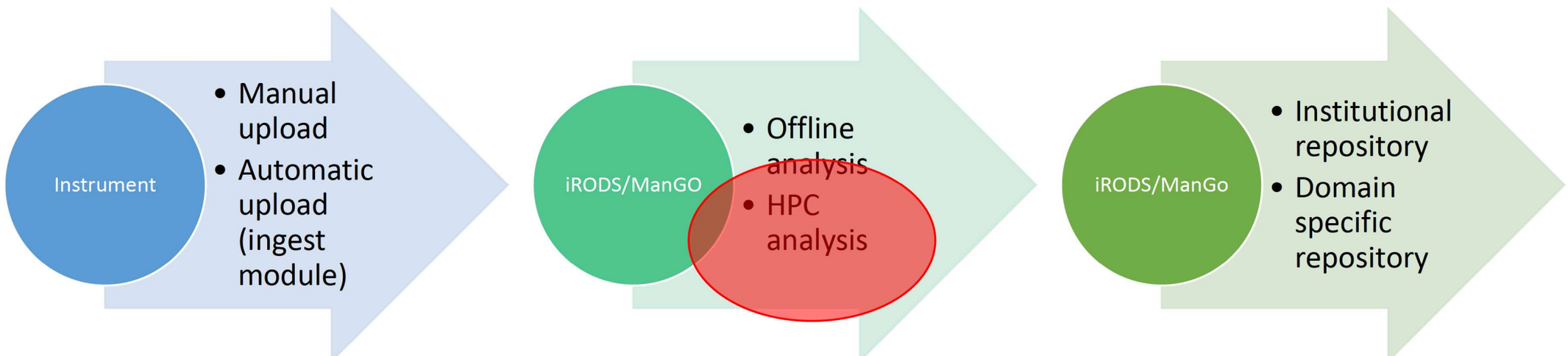
- Info on file reader
- .nd2-reader (Nikon file format)
- Work in progres

```
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
```

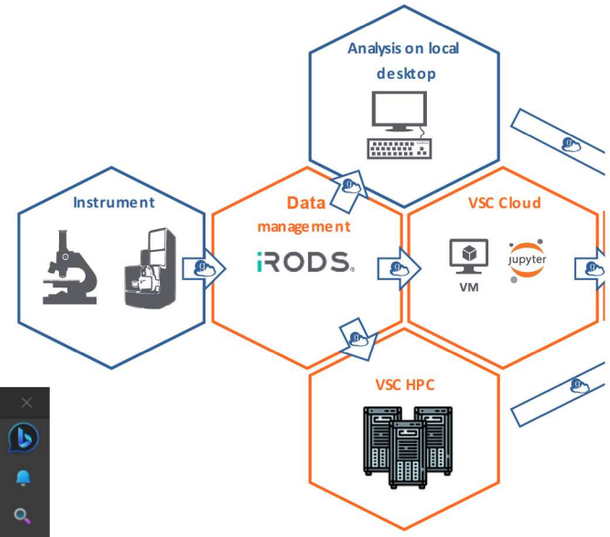
✓ 0.8s Python

	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)

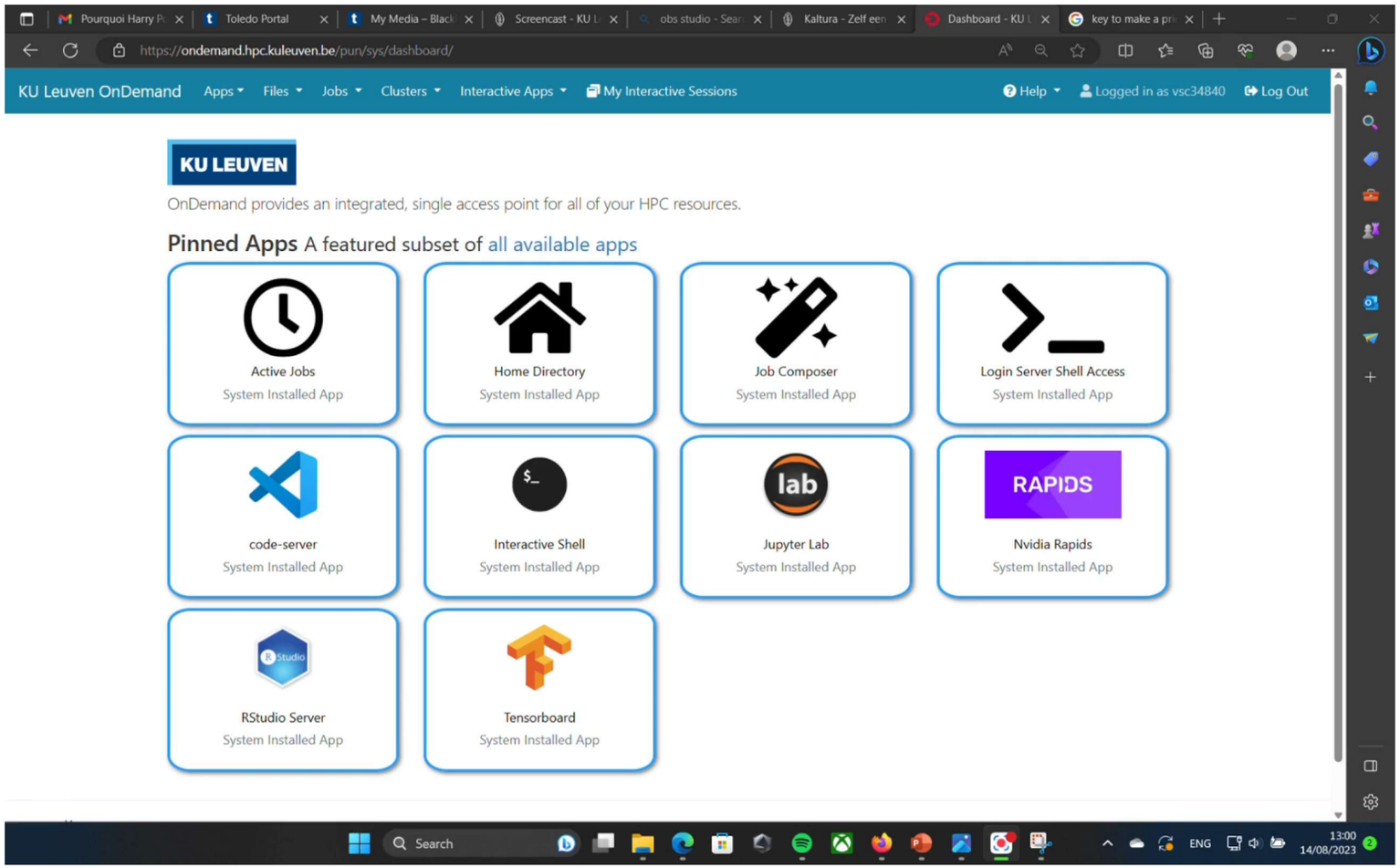
Now that we uploaded the data – Connecting with analysis



Analysis using Open-OnDemand

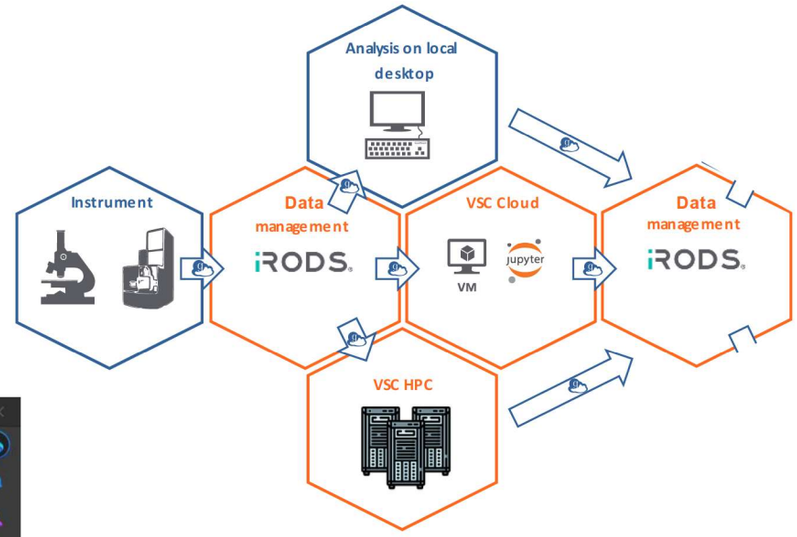


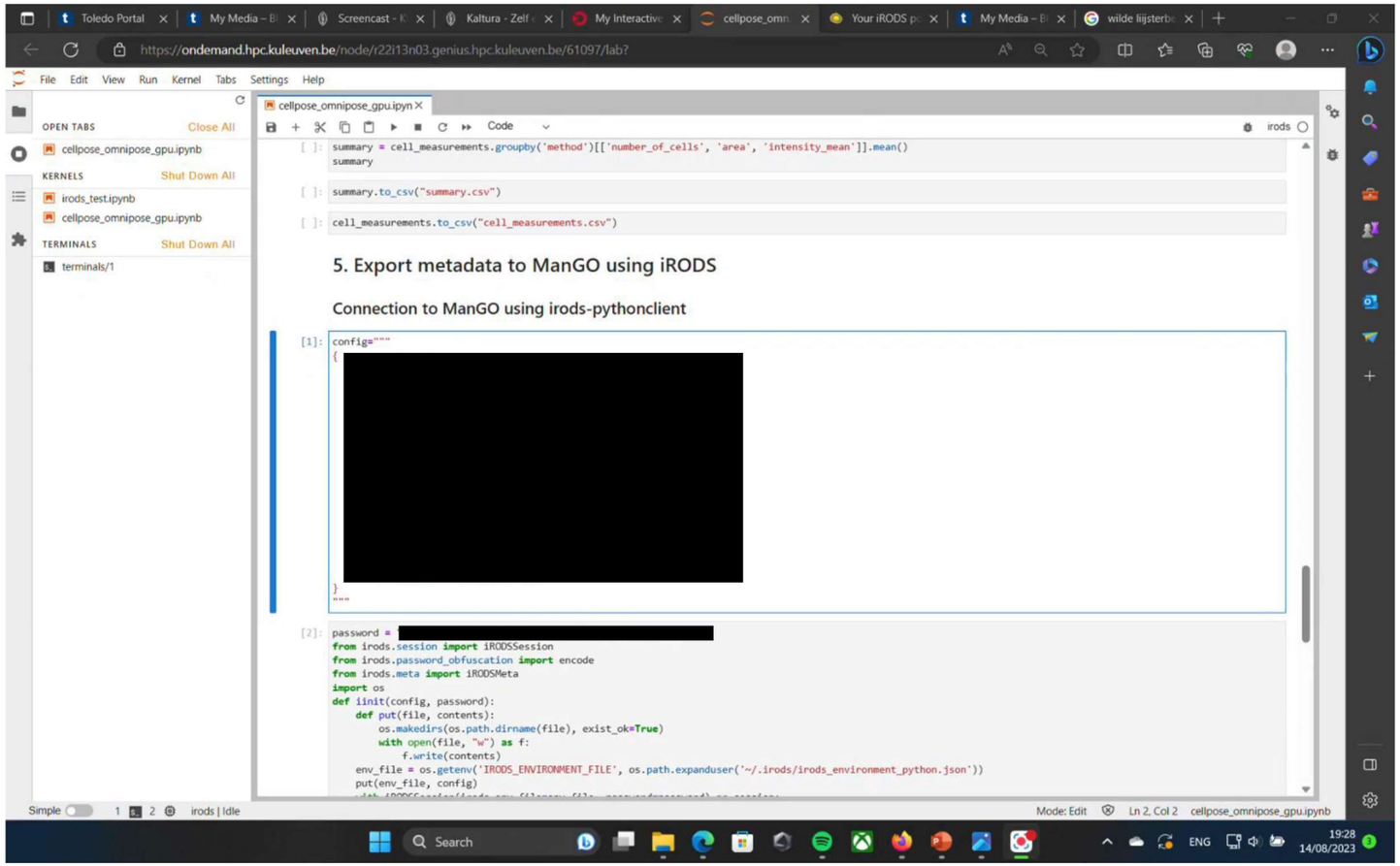
Till Korten & Tatiana



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

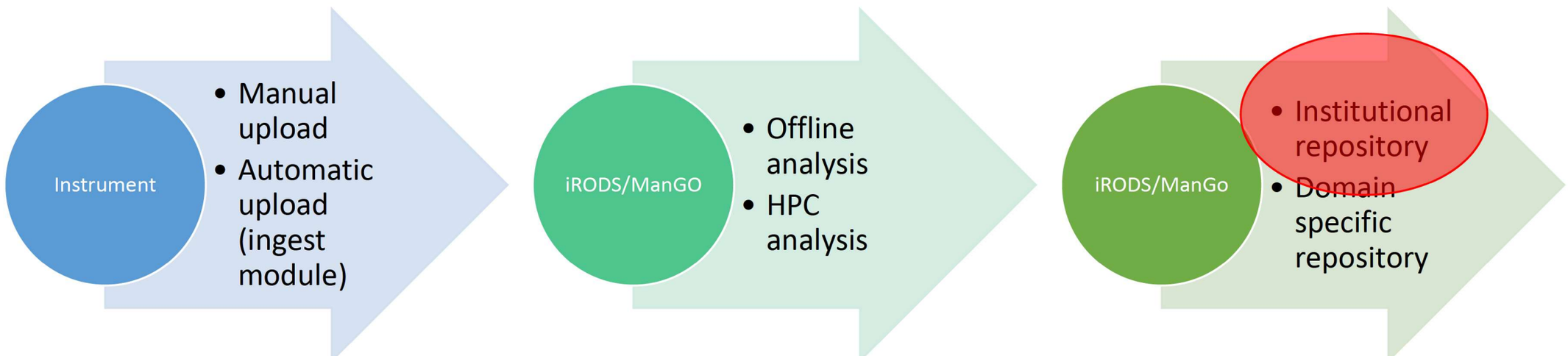
From the analysis to ManGO





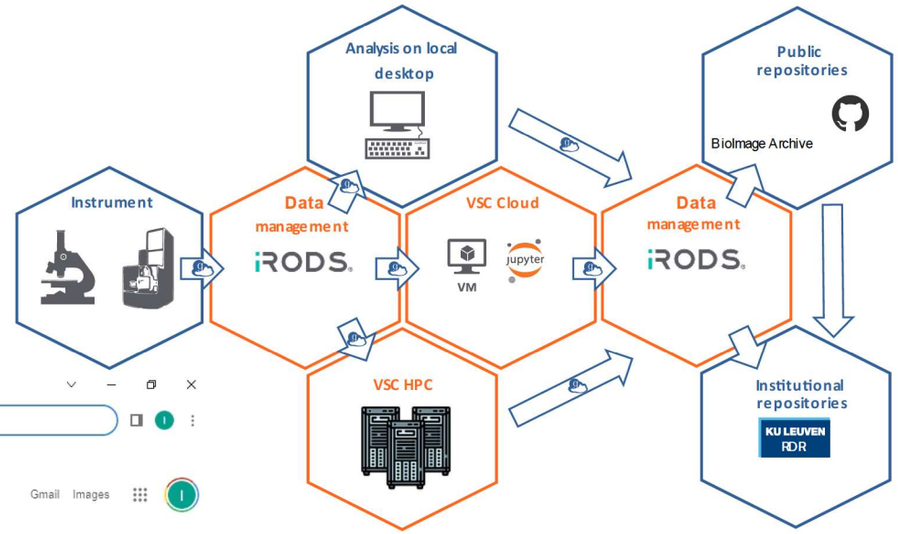
- Connecting to ManGO using a python irods-client.
- Checking that the file to which we add metadata is present
- Create a new folder, “analysis,” in ManGO and export the csv file together with some metadata.
- On the ManGO platform checking whether the csv file is present and if it has the right metadata.

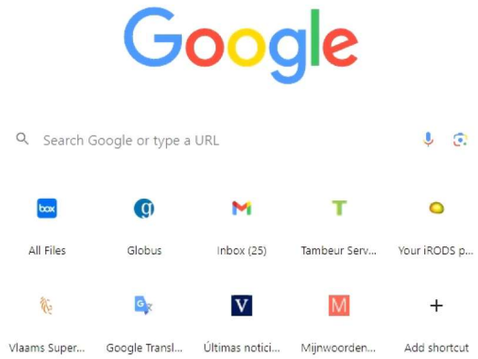
Publicizing the data



Publish to an institutional repository

proof of concept

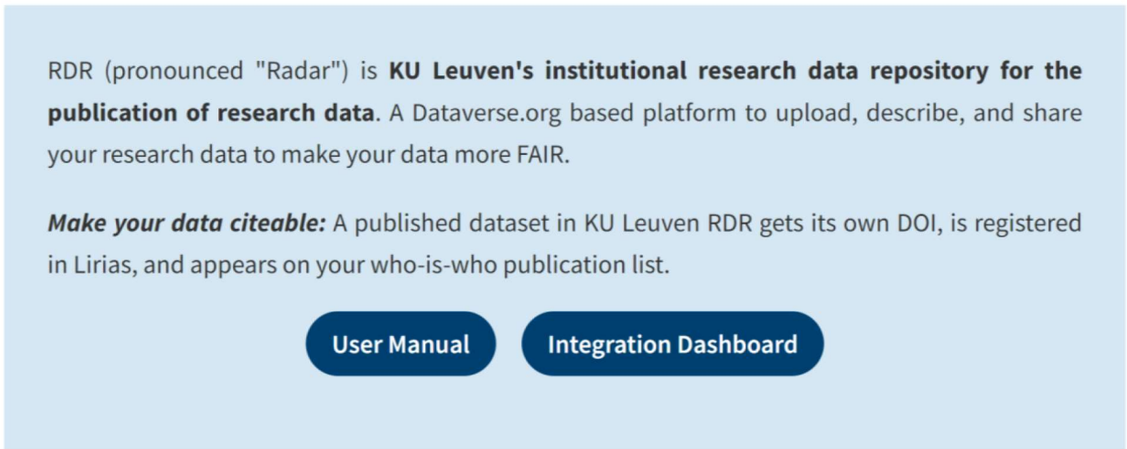




- Create a dataset on RDR (the KU Leuven data repository)
- Use the RDR Integration tool to directly select and upload files to RDR

Local repository to publish research data

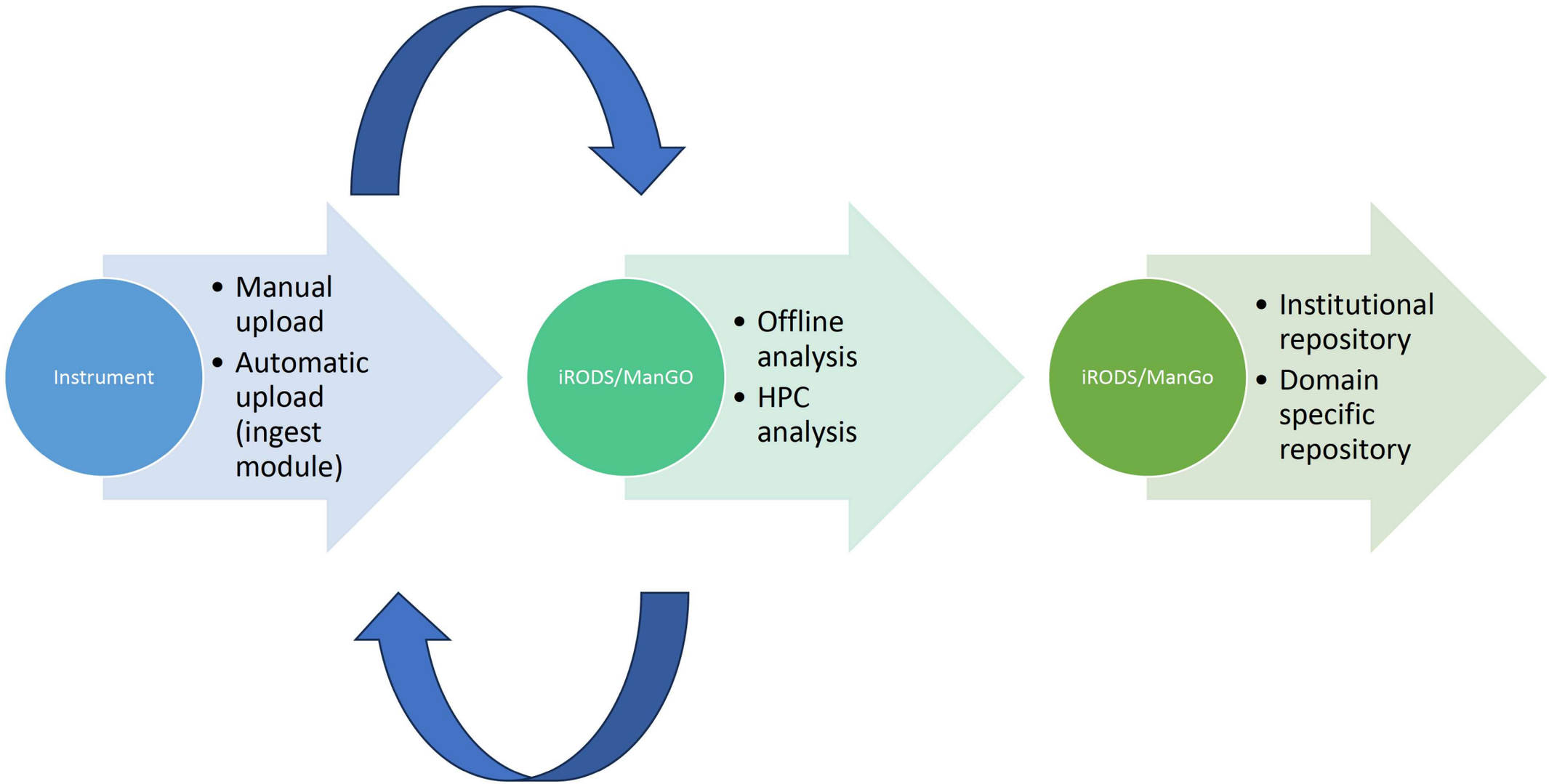
KU Leuven RDR



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

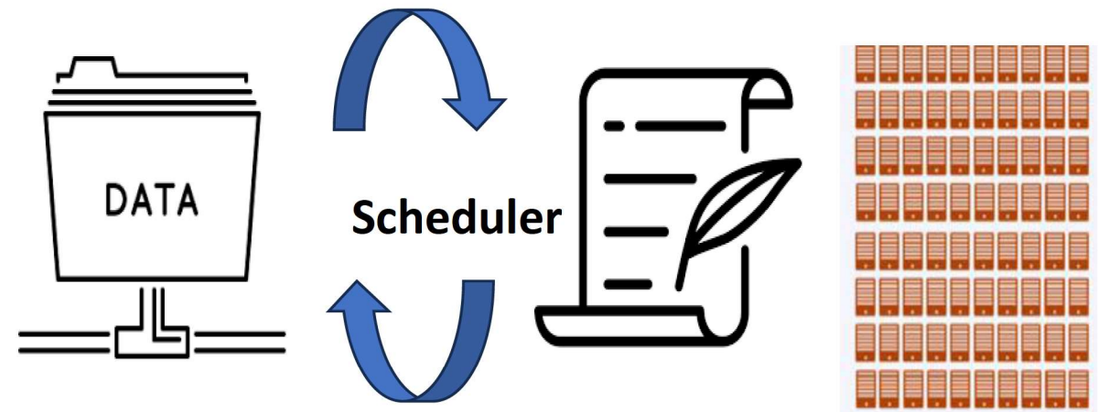
Individual analysis (like in these jupyter notebooks) is great – but how about automating part of the processing in a 'blind' fashion?

Automating more steps in the workflow



SCROn-jobs: Work in progress

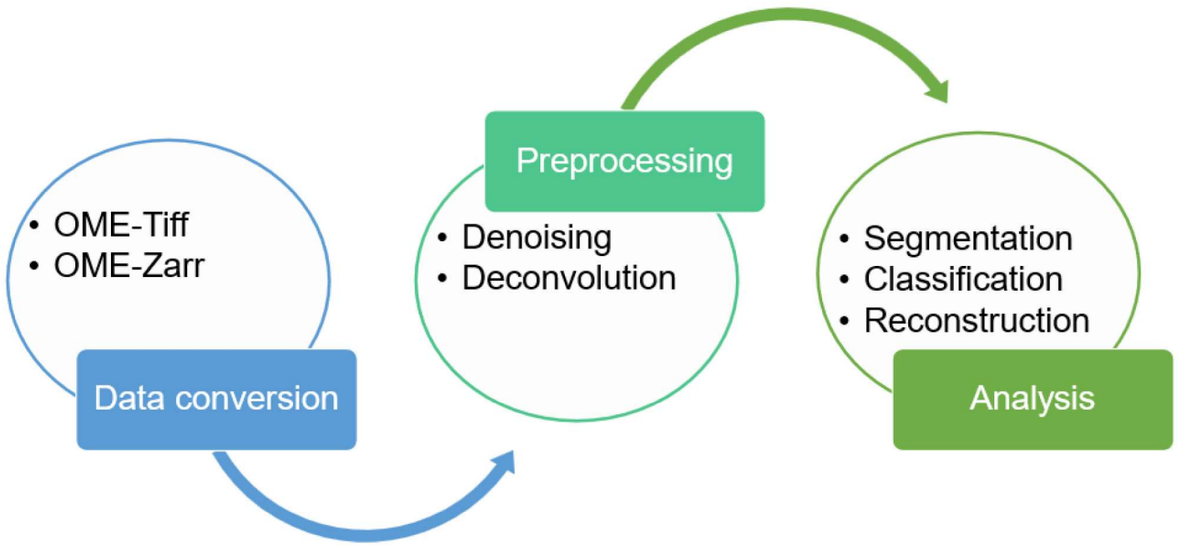
- A **cron job** is a Linux command used for scheduling tasks to be executed sometime in the future. This is normally used to schedule a job that is executed periodically – for example, to send out a notice every morning. Scron joba are targeted at Slurm clusters



Ingest module submitting data + job token

Processing script

Slurmjob



Potential automatic jobs

Automatic ingestion

Format conversion metadata schema:

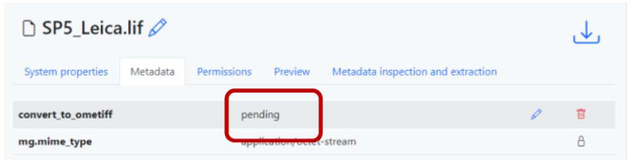
- To OME-TIFF: pending|done
- To OME-ZARR: pending| done

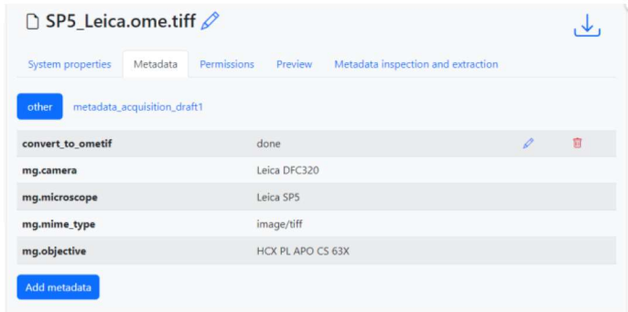
.lif files

Scronjob -daily

The job will:

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

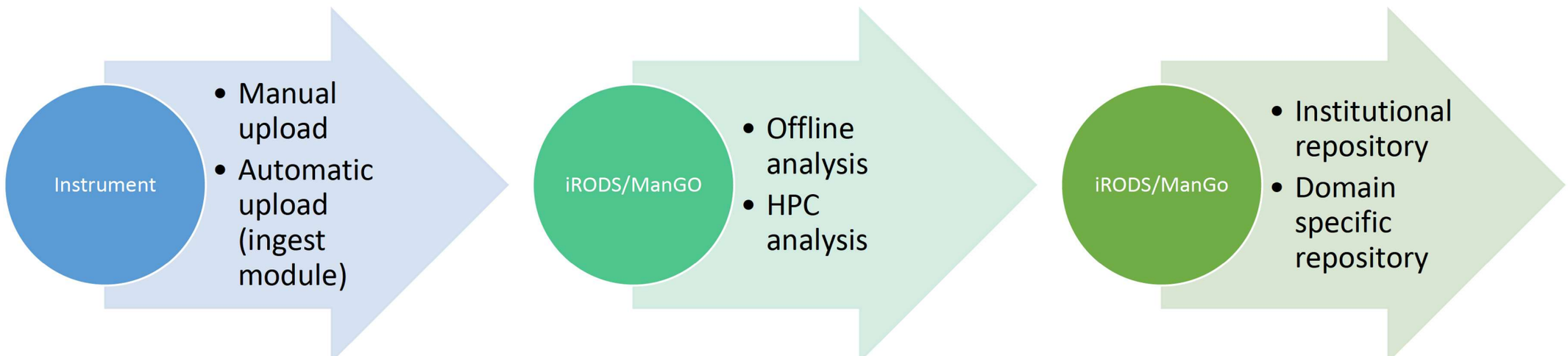




.tiff files

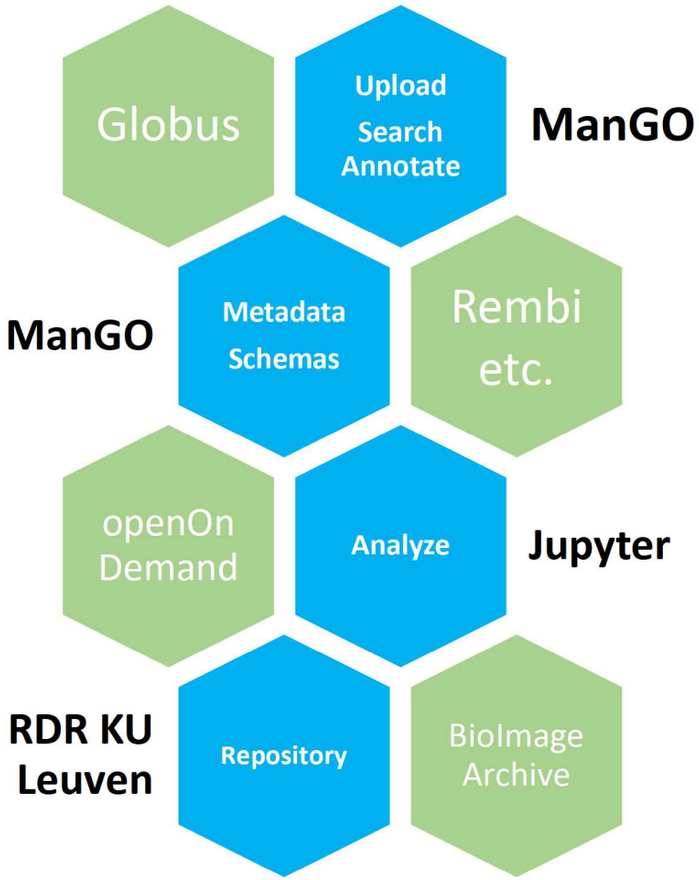
BatchConvert

A managed workflow



Quick summary: The data journey through connected tools

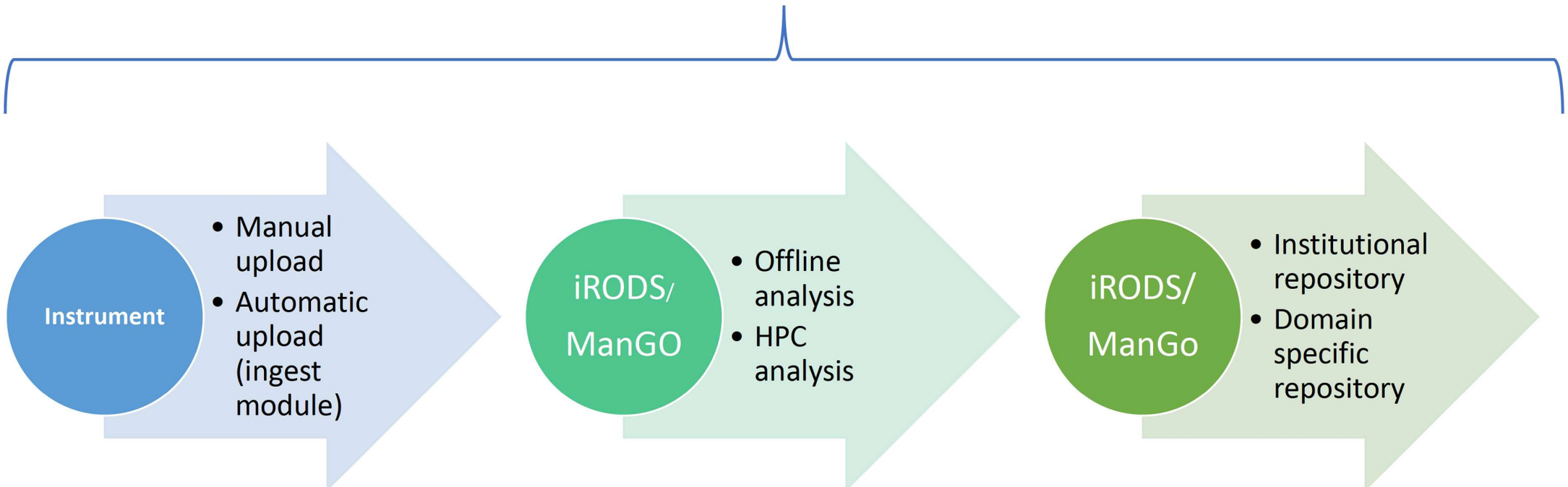
FAIR by design



- 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

ELN

LIMS

Acquisition

ZEISS · Nikon · Leica MICROSYSTEMS

XNAT

OME

OMERO

Analysis

GitHub

Galaxy PROJECT

?

Data Management

iRODS

ManGO

NFS NETWORK FILE SYSTEM

Some REMBIsh combining information

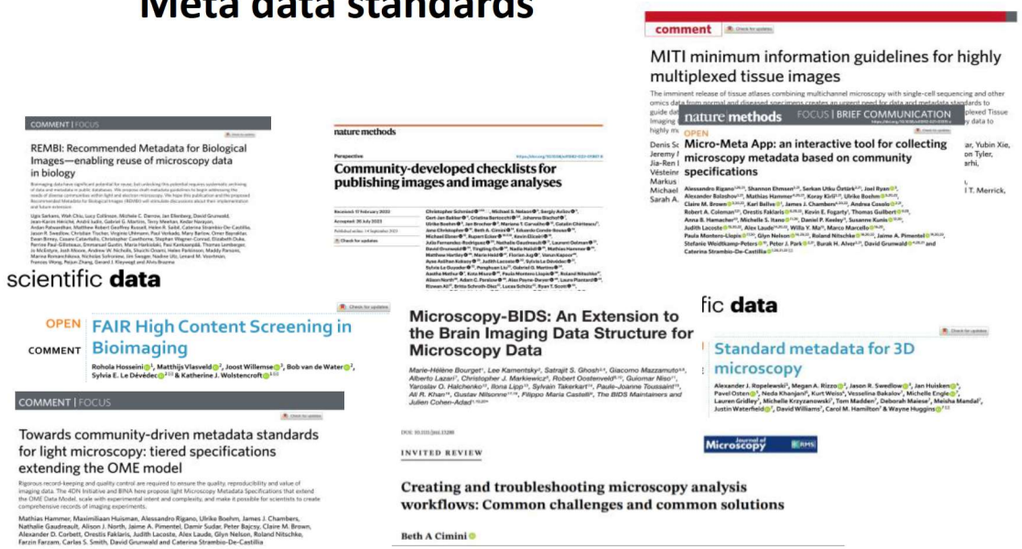
BioImage Archive

Other ontologies

KU LEUVEN RDR RESEARCH DATA REPOSITORY

EMPIAR Electron Microscopy Public Image Archive

Meta data standards



My Electronic lab-notebook

Initiatives, stakeholders, and Funders

Institute policies

BBMRI-ERIC

INFRAFRONTIER

elixir

eatris European infrastructure for translational medicine

BioImage Archive

EURO-BIOIMAGING Image life, discover the future

eu-openscreen

instruct ERIC

AnaEE ERIC Analysis and Experimentation on Ecosystems

incf

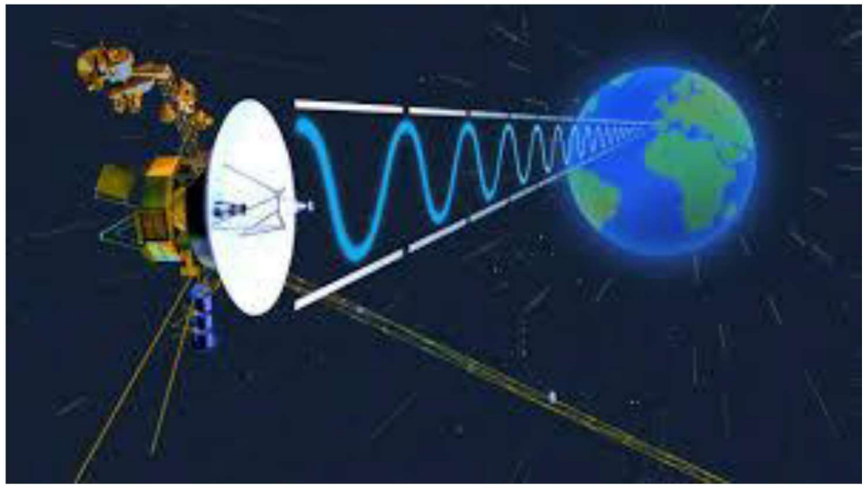
...

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

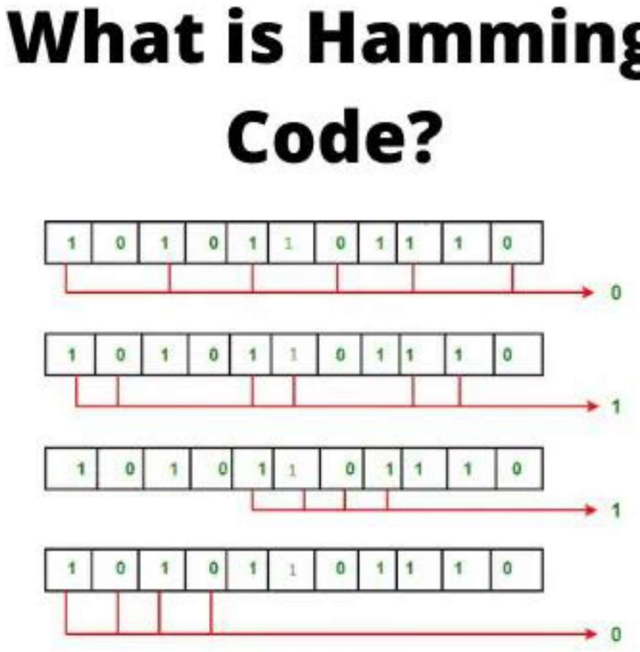
…you?

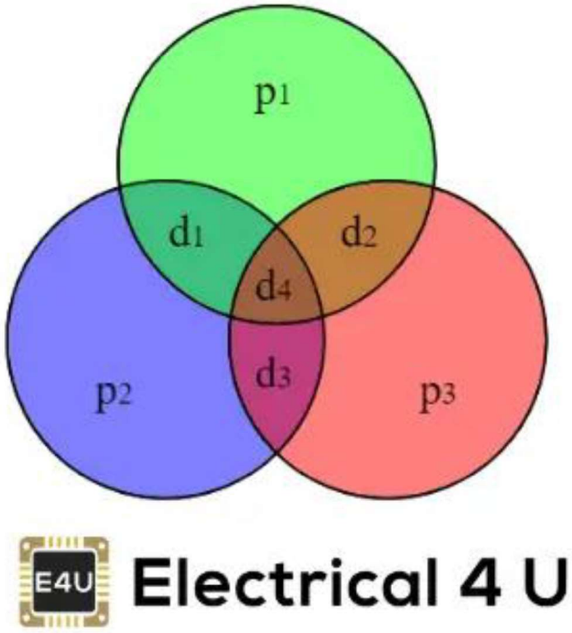
How do we combine the *scattered* information and correct errors?

Maybe we can use the redundancy to our advantage



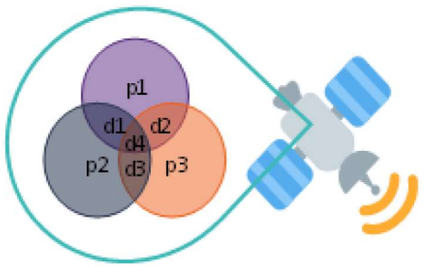
Communication is also difficult in other disciplines



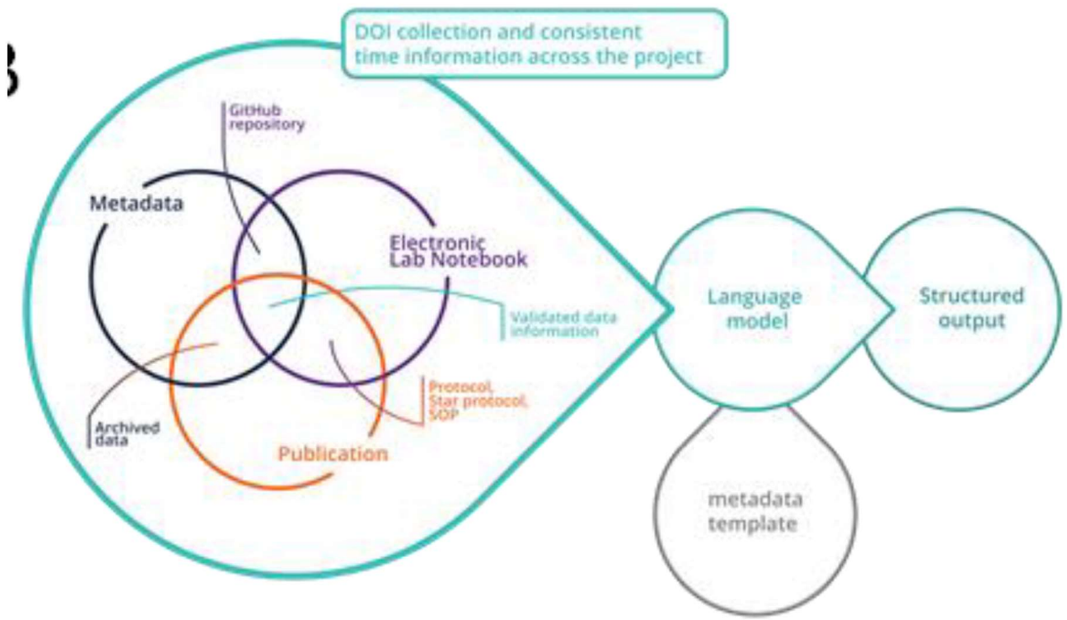


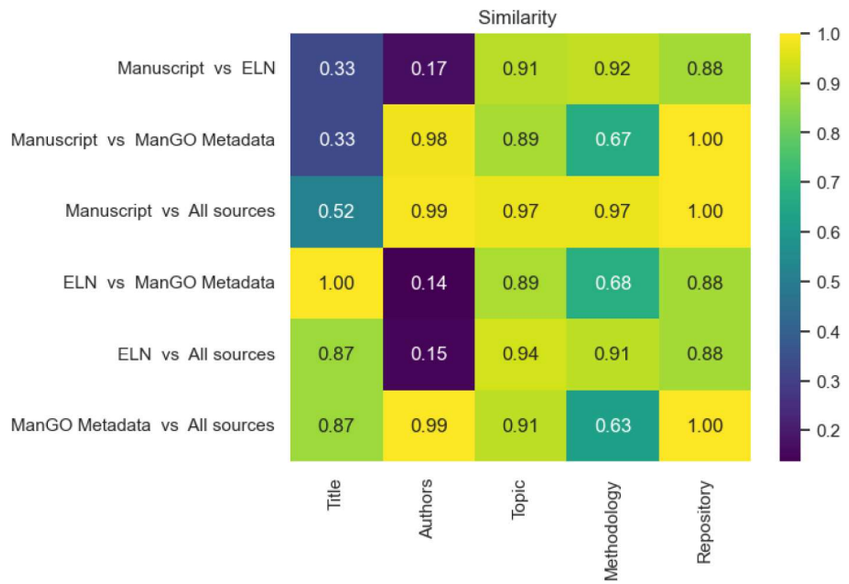
Data can be transported over noisy channels (Shannon), and it can be corrected (Hamming)

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





Woller et al. Mol. Systems Biology. 2023

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

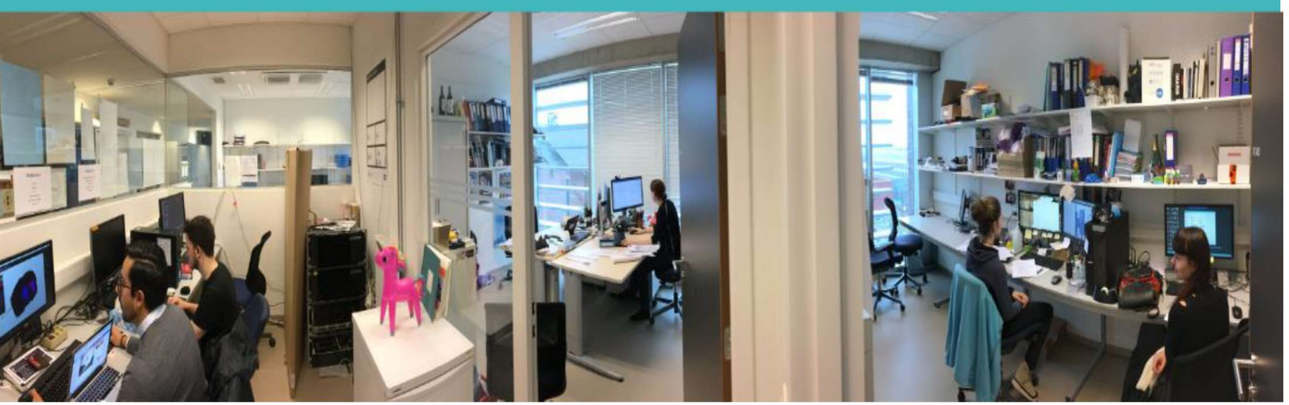
find me @

VIB-KU LEUVEN

CENTER FOR BRAIN & DISEASE RESEARCH

cbd.vib.be

@BioImagingCore; @SebastianMunck



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

VIB Technology Training
Alex Botzki
Tatiana Woller

TU Dresden
Till Korten
Robert Haase

University of Antwerp
Winnok De Vos
Marlies Verschuuren
Tim Van De Looverbosch

KU Leuven Nuclear medicine & molecular imaging
Koen Van Laere
Chris Cawthorne

VSC
Ingrid Barcena-Roig
Jan Ooghe
Mariana Montes