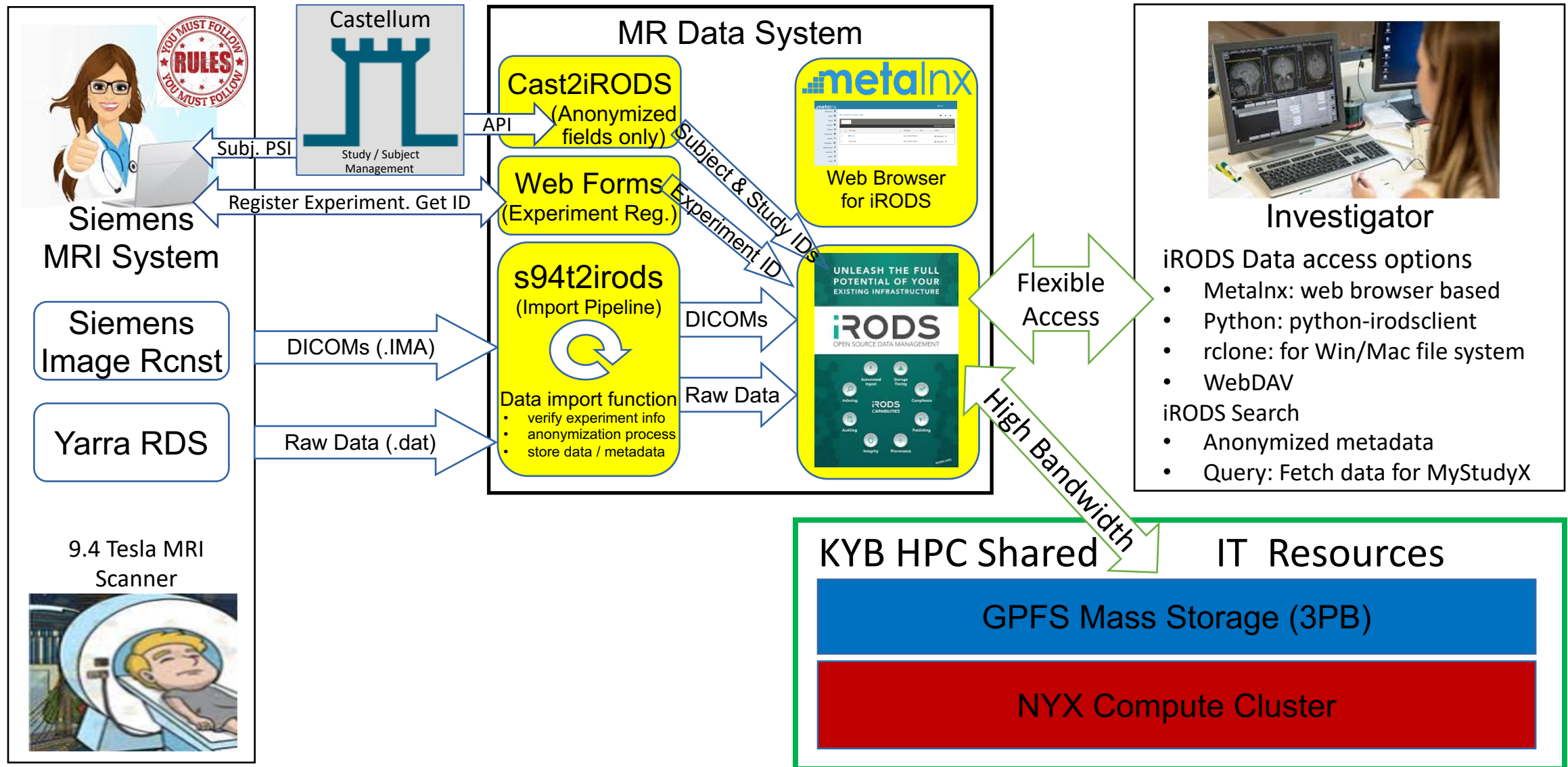


MrData: An iRODS Based Human Research Data Management System

iRODS User Group Meeting
6 July 2022

Blake G. Fitch
blake.fitch@tuebingen.mpg.de

The MrData Project for Human Research Data Archival



Overview

Castellum is a human subject database system developed at the Max Planck Society. Its main goals are:

- **GDPR** compliant data protection and security
- Flexibility so it can be used in different organizations

Features

Subject management

- Castellum is a central place to collect references to all data related to a subject, e.g. so it can be deleted on request

Pseudonym service

- Contact details are stored in Castellum so all other databases can work with pseudonyms instead.

Recruitment

- Castellum allows you to find potential subjects from an existing pool using study specific filters

Appointments

- You can manage appointments for experiment sessions

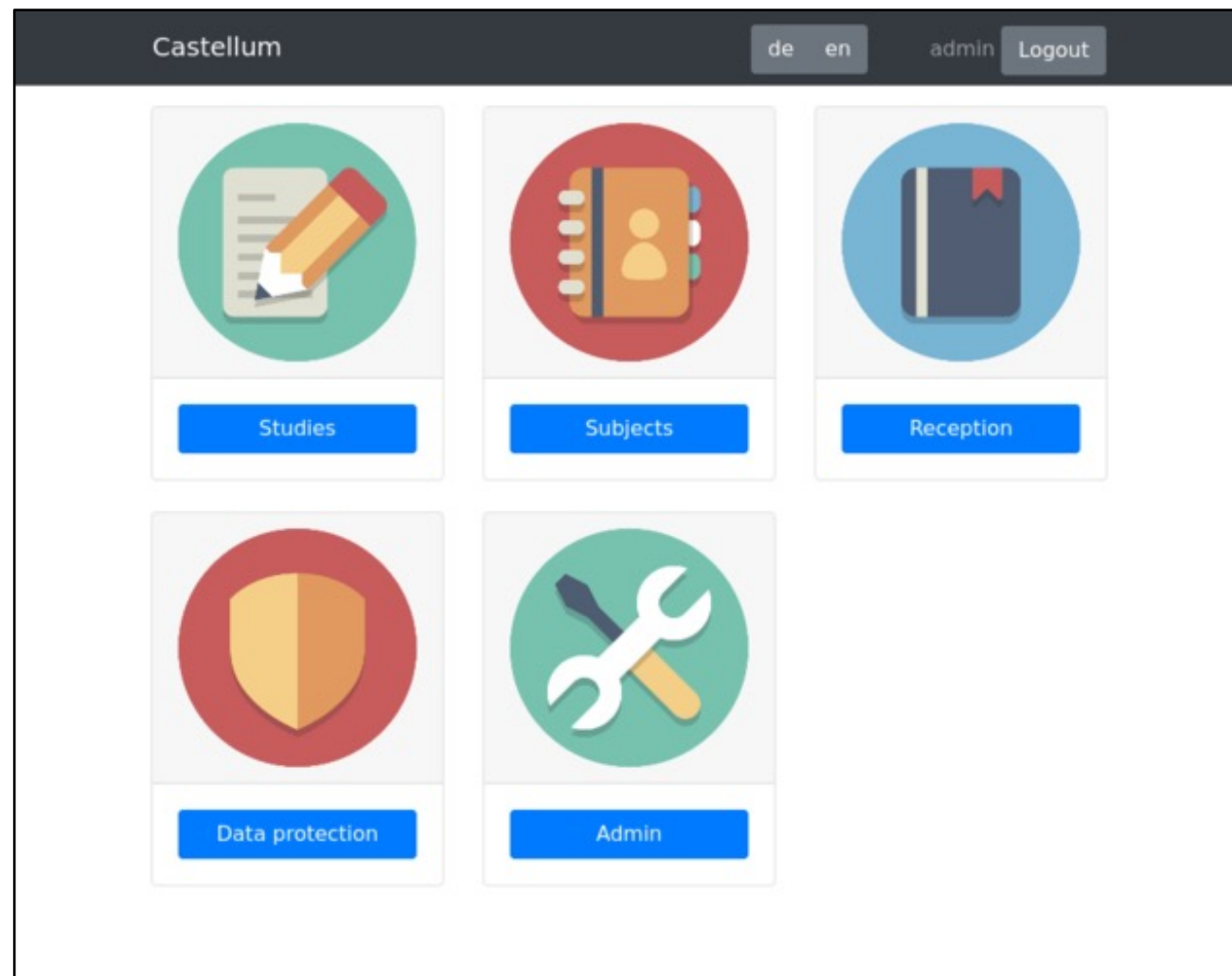
Application Programming Interface (API) (optional)

- Available API for exportable attributes. (pseudonyms, etc) which are needed to integrate with a data management system

What Castellum is not

Castellum does not store scientific data.

It manages information about studies and subjects required for recruiting and provides subject pseudonyms for external use.



<https://castellum.mpib.berlin/documentation/en/overview.html>

Mixed Use Metadata: A Challenge For Human Research

Certain bits of metadata are required for both:

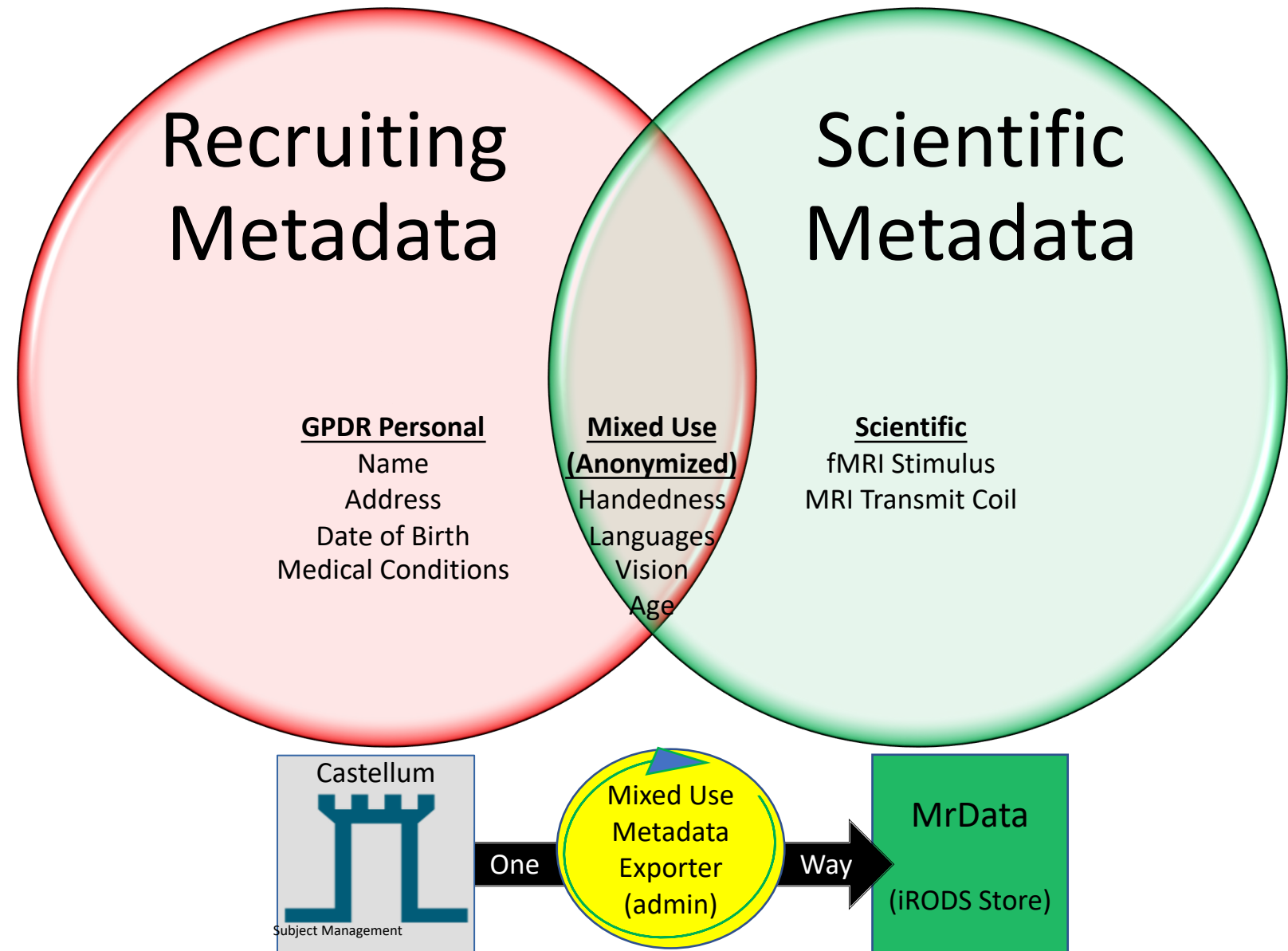
- recruiting activities
- scientific search and catalog

Rules for Mixed Use Metadata:

- GDPR Personal metadata can only be kept in Castellum
- A given bit of metadata should be acquired from a single, authoritative source
- Mixed use metadata can only be sourced from Castellum

Selected, GDPR safe, subject attributes move one-way from Castellum to MrData by admin domain automation.

Uses Castellum API for safe data access.



<https://www.accountablehq.com/post/pii-vs-phi>

MrData Project -- Blake G. Fitch -- MPI KYB

MrData Experiment Registration Ties It All Together

Experiment Registration

- Associates StudyID and Subject pseudonym with the ExperimentID
- The ExperimentID is entered in place of a "Patient Name" on the MRI scanner
- This avoids contaminating MRI scanner data files with personal information about the subject...
- but does require maintaining this information for the life of the data
- This association provides search mixed use metadata
- An admin operation can remove all data for a subject

Experiment Registration Web Pages: Request/Response

Register an Experiment

Experiment Owner (campus user id)

Subject Psuedonym (from Castellum)

Study ID (from Castellum)

Scanner

Siemens 9.4T

Scan Type

Human Scan

Experiment URL

Experiment Description (NO GDPR VIOLATIONS!)

Submit

Form Capture Confirmation

Form

Experiment Registration

New Experiment ID (Save! Needed at Scanner!)

Z4KQ-STAP [Copy to Clipboard](#)

Form Data (yaml)

```
---
Description: "
ExperimentID: Z4KQ-STAP"
```

[Go To Main Page](#)

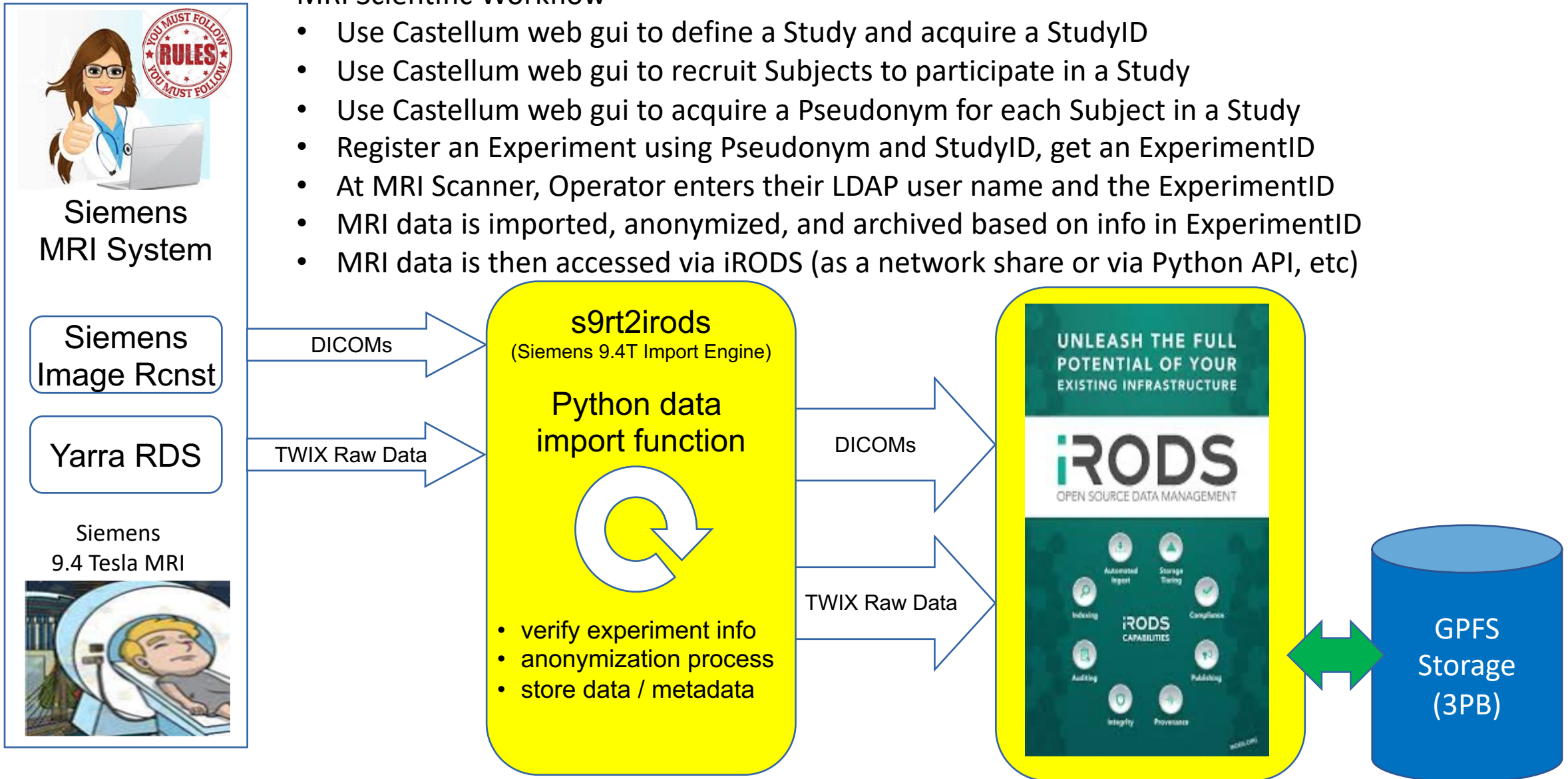
These two web pages are implemented as a small Python Flask app in the Forms container.

Note: Prior to using Castellum, other web pages for Subject and Study registration existed.

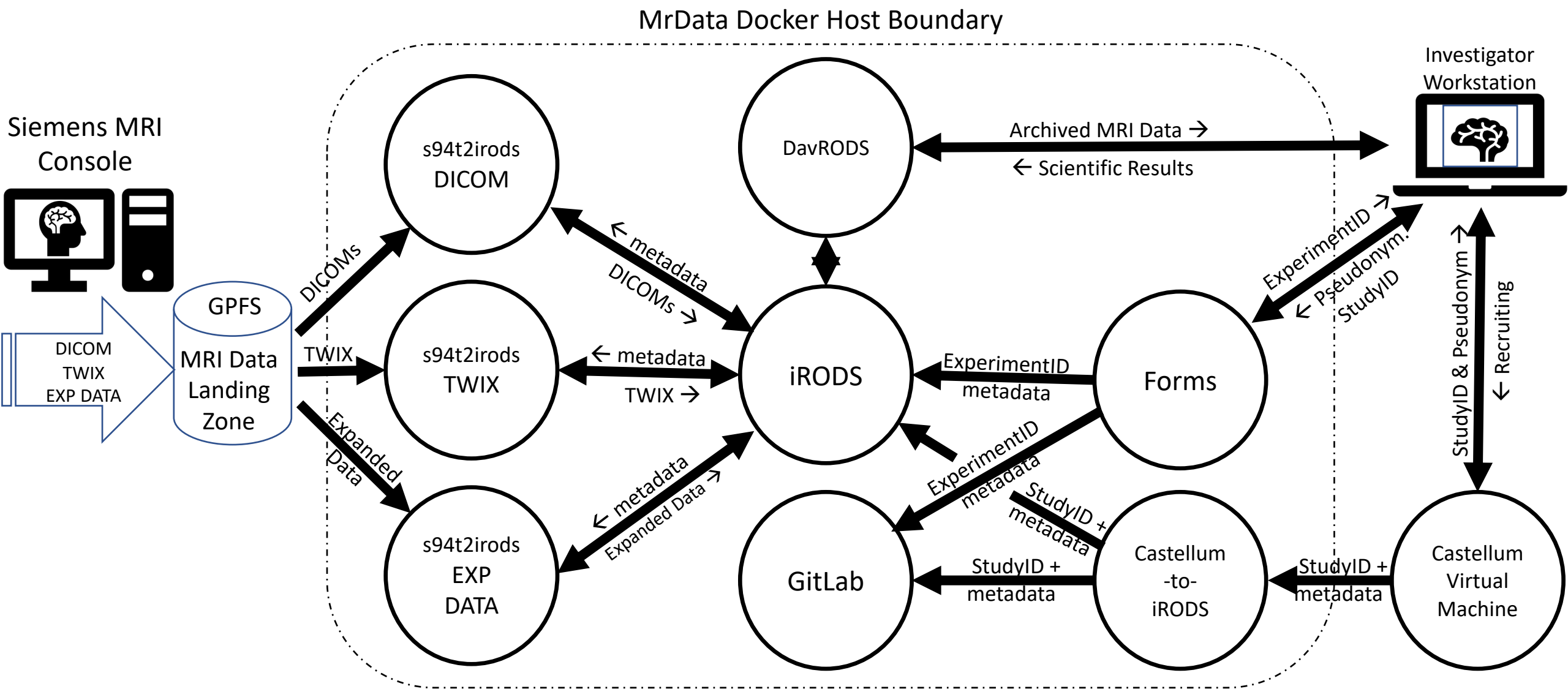
MrData MRI Research Investigator Workflow

MRI Scientific Workflow

- Use Castellum web gui to define a Study and acquire a StudyID
- Use Castellum web gui to recruit Subjects to participate in a Study
- Use Castellum web gui to acquire a Pseudonym for each Subject in a Study
- Register an Experiment using Pseudonym and StudyID, get an ExperimentID
- At MRI Scanner, Operator enters their LDAP user name and the ExperimentID
- MRI data is imported, anonymized, and archived based on info in ExperimentID
- MRI data is then accessed via iRODS (as a network share or via Python API, etc)



MrData: Infrastructure Environment and Microservice App

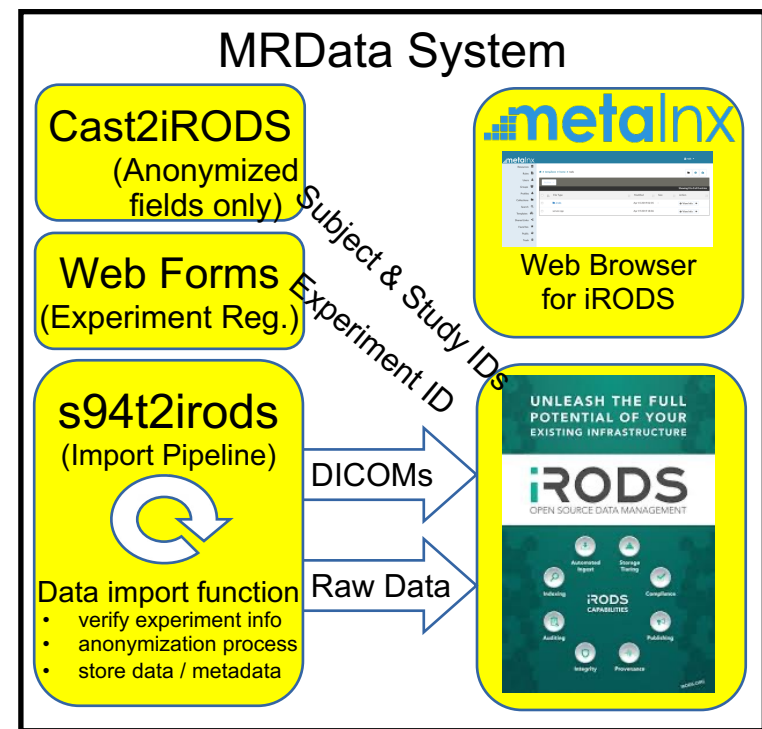


MrData Component Services Live in Separate Containers

```
mradmin@mrdata01:~  
File Edit View Terminal Tabs Help  
Every 1.0s: docker ps  
CONTAINER ID   IMAGE                                COMMAND                                  CREATED        STATUS        PORTS  
a8c5f1551e51   s94t2irods_image_prod              "/home/mradmin/mr2ir..."             3 weeks ago   Up 3 weeks   (healthy)     
c5db68e321fa   s94t2irods_image_prod              "/home/mradmin/mr2ir..."             3 weeks ago   Up 3 weeks   (healthy)     
9d64cce8bdca   s94t2irods_image_prod              "/home/mradmin/mr2ir..."             3 weeks ago   Up 2 weeks   (healthy)     
daeb64bb9a0e   cast2irods_image_prod              "/home/mradmin/cast2..."             3 weeks ago   Up 3 weeks     
32e3a370ec28   forms_image_prod                   "/home/mradmin/mrfor..."             4 weeks ago   Up 4 weeks   8080/tcp  
2bffcc4498d4   nginx:latest                       "/docker-entrypoint..."             7 weeks ago   Up 7 weeks   0.0.0.0:80->80/tcp, 0.0.0.0:443->443/tcp  
b876fc90054b   davrods_image_prod                 "/bin/sh -c 'dockeri..."             7 weeks ago   Up 7 weeks   80/tcp  
7612260e1508   irods_image_prod                   "/docker-entrypoint..."             7 weeks ago   Up 7 weeks   0.0.0.0:1247-1248->1247-1248/tcp, 0.0.0.0:5432->5432/tcp, 0.0.0.0:20000-20199->20000-20199/tcp  
NAMES  
s94t2irods_EXP_DATA_prod  
s94t2irods_TWIX_prod  
s94t2irods_DICOM_prod  
cast2irods_prod  
forms_prod  
nginx  
davrods_prod  
irods_prod
```

This is a Docker based micro-services based architecture

- Pros
 - Integrates home grown and external services
 - Single Docker host (currently)
 - Ansible deployed docker containers for all services
 - More robust than Docker Compose, easier than K8s, etc
 - Possible to test and redeploy individual services (via Ansible)
 - Independent test infrastructure can be deployed to Vbox VM
 - Full system CI/CD possible though not nearly completed
 - Extensible within the Docker ecosystem
- Cons
 - micro-services, docker, ansible – the usual stuff



Data layout in the iRODS Data “VAULT” on Linux FS

iRODS internal path:

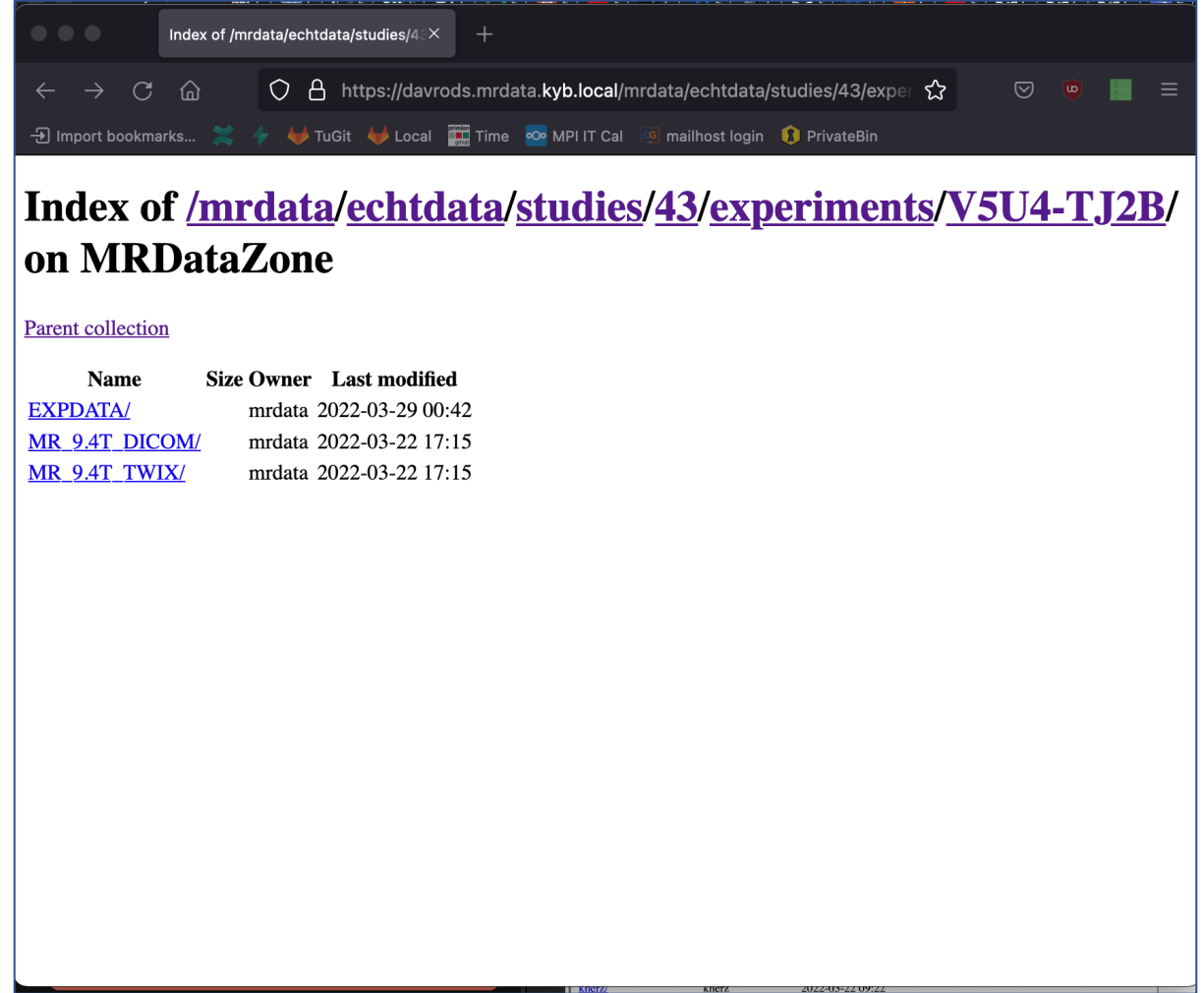
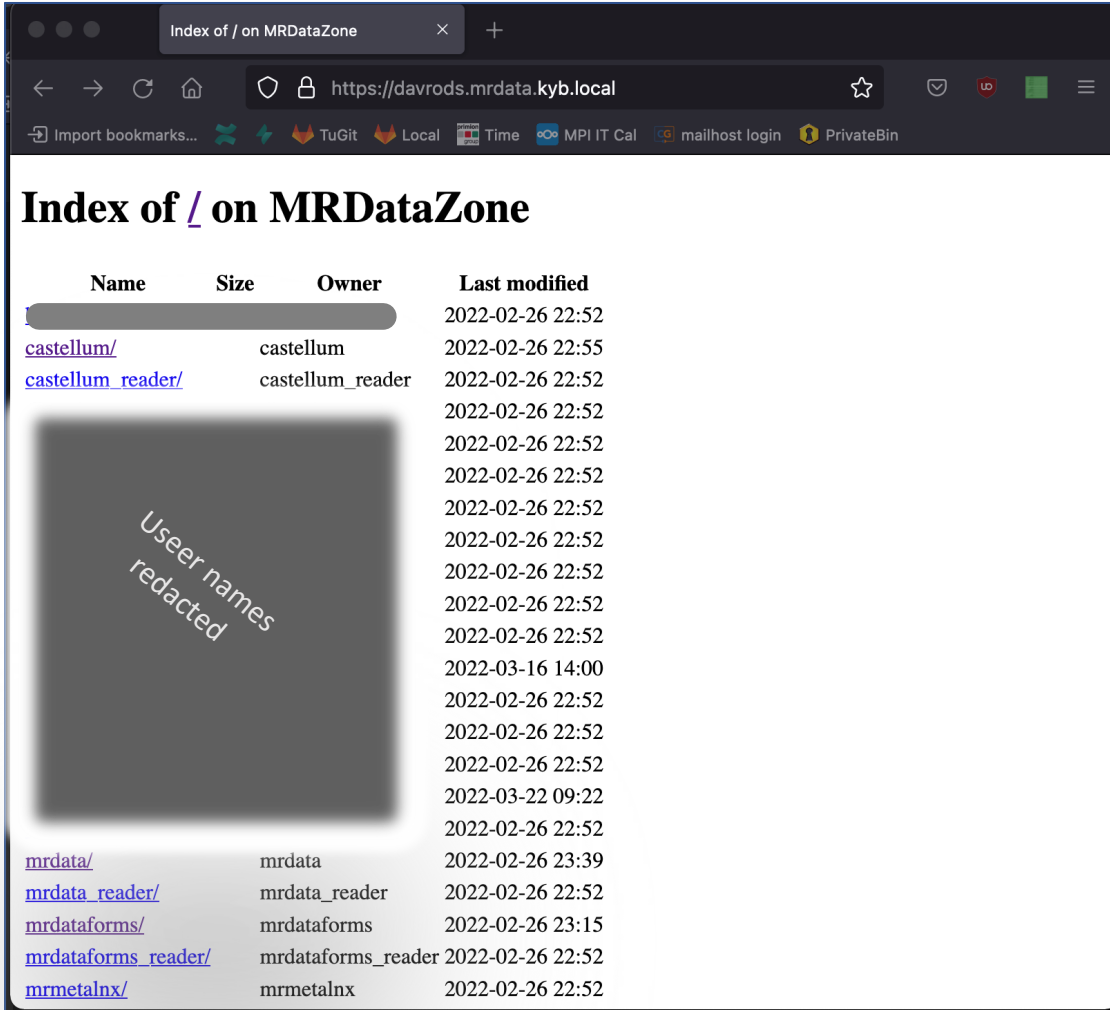
/MRDataZone/home/mrdata/echtdata/studies/<StudyID>/experiments/<ExperimentID>/<data_type>/

```
[mradmin@mrdata01 ~]$ sudo find /zfs-pool/MRDATA/prod/irods/IRODS_VAULT/home/mrdata/echtdata/studies/43/experiments -maxdepth 2
/zfs-pool/MRDATA/prod/irods/IRODS_VAULT/home/mrdata/echtdata/studies/43/experiments
/zfs-pool/MRDATA/prod/irods/IRODS_VAULT/home/mrdata/echtdata/studies/43/experiments/HGZD-0037
/zfs-pool/MRDATA/prod/irods/IRODS_VAULT/home/mrdata/echtdata/studies/43/experiments/HGZD-0037/MR_9.4T_TWIX
/zfs-pool/MRDATA/prod/irods/IRODS_VAULT/home/mrdata/echtdata/studies/43/experiments/HGZD-0037/EXPDATA
/zfs-pool/MRDATA/prod/irods/IRODS_VAULT/home/mrdata/echtdata/studies/43/experiments/HGZD-0037/MR_9.4T_DICOM
/zfs-pool/MRDATA/prod/irods/IRODS_VAULT/home/mrdata/echtdata/studies/43/experiments/V5U4-TJ2B
/zfs-pool/MRDATA/prod/irods/IRODS_VAULT/home/mrdata/echtdata/studies/43/experiments/V5U4-TJ2B/EXPDATA
/zfs-pool/MRDATA/prod/irods/IRODS_VAULT/home/mrdata/echtdata/studies/43/experiments/V5U4-TJ2B/MR_9.4T_TWIX
/zfs-pool/MRDATA/prod/irods/IRODS_VAULT/home/mrdata/echtdata/studies/43/experiments/V5U4-TJ2B/MR_9.4T_DICOM
[mradmin@mrdata01 ~]$
```

Questions?

WebDav Interface: See what's in iRODS via browser login

iRODS internal 'ipath': /MRDataZone/home/mrdata/echtdata/studies/<StudyID>/experiments/<ExperimentID>/



MrData view from MacOS Finder (iRODS mounted as a file system)

