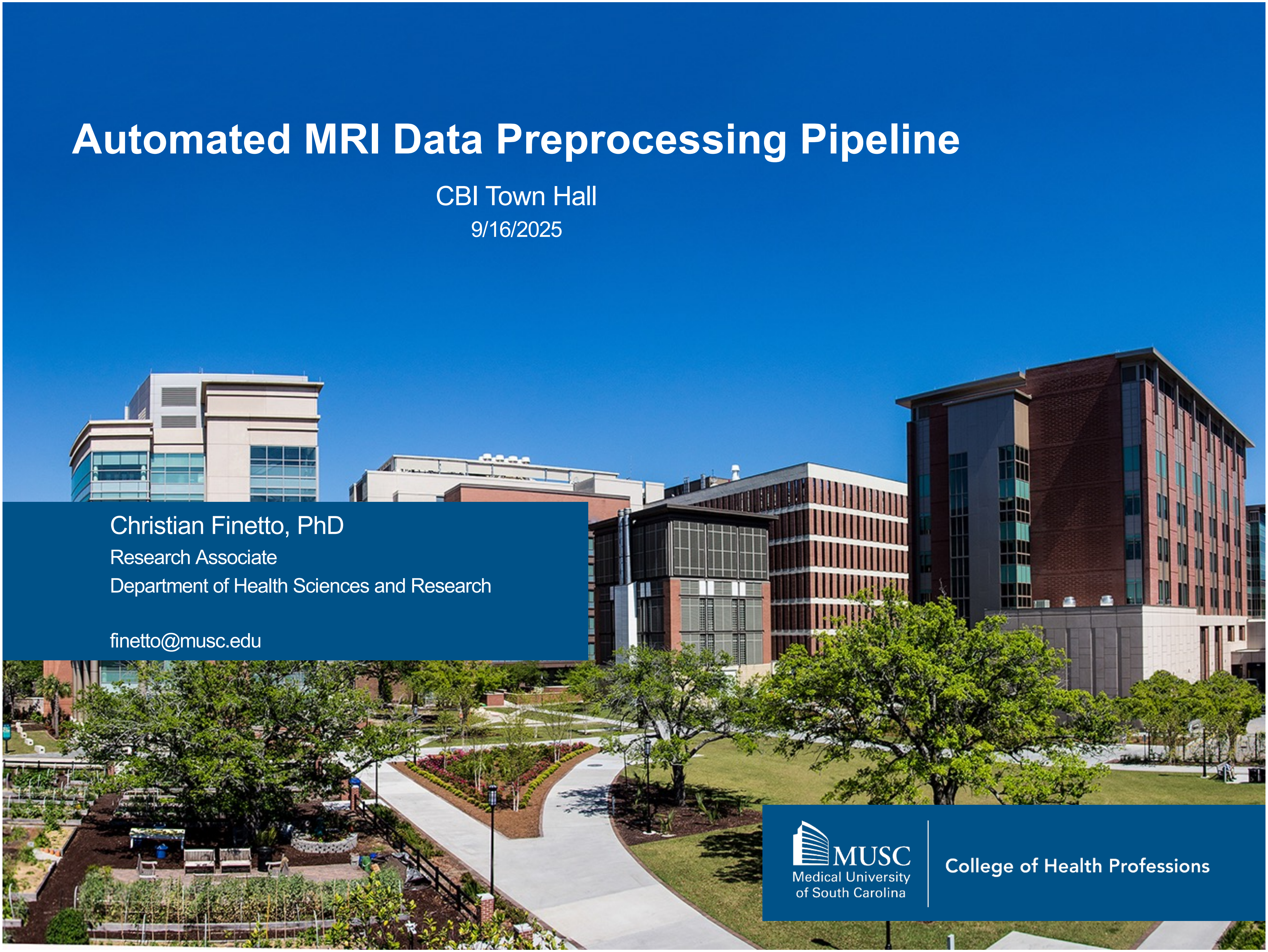


Automated MRI Data Preprocessing Pipeline

CBI Town Hall

9/16/2025

Christian Finetto, PhD
Research Associate
Department of Health Sciences and Research

finetto@musc.edu

MUSC Medical University of South Carolina | College of Health Professions

Automated MRI Data Preprocessing Pipeline

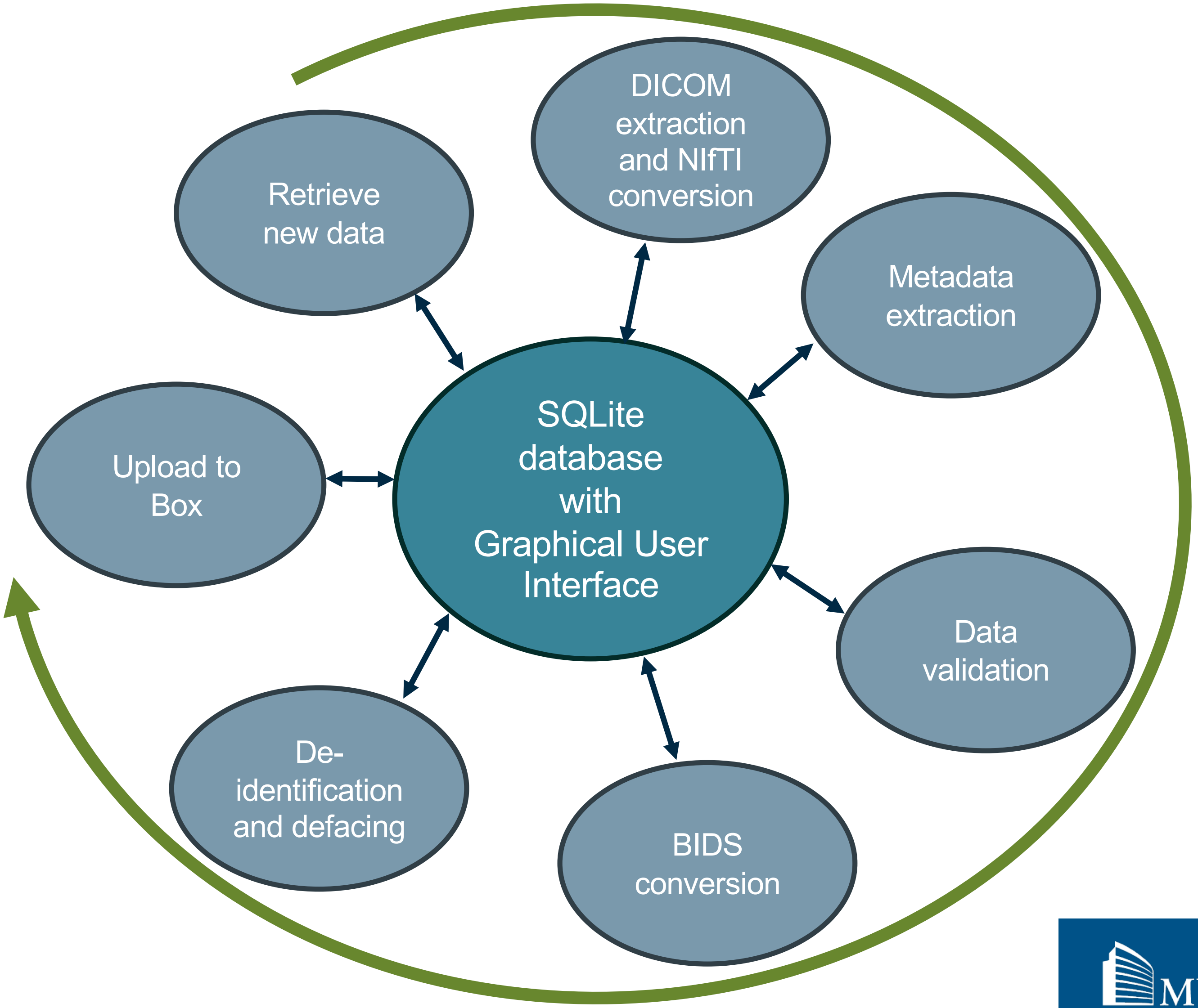
We developed this pipeline for Dr. Lisa McTeague’s DARPA RECOVERS study to streamline data preparation and automate repetitive tasks:

- Data retrieval (from CBI Home or from local folder)
- DICOM extraction and NIfTI conversion
- Metadata extraction
- Data validation
- Conversion to BIDS format
- De-identification and defacing
- Upload to Box

The pipeline can be easily adapted to other studies by modifying the configuration files

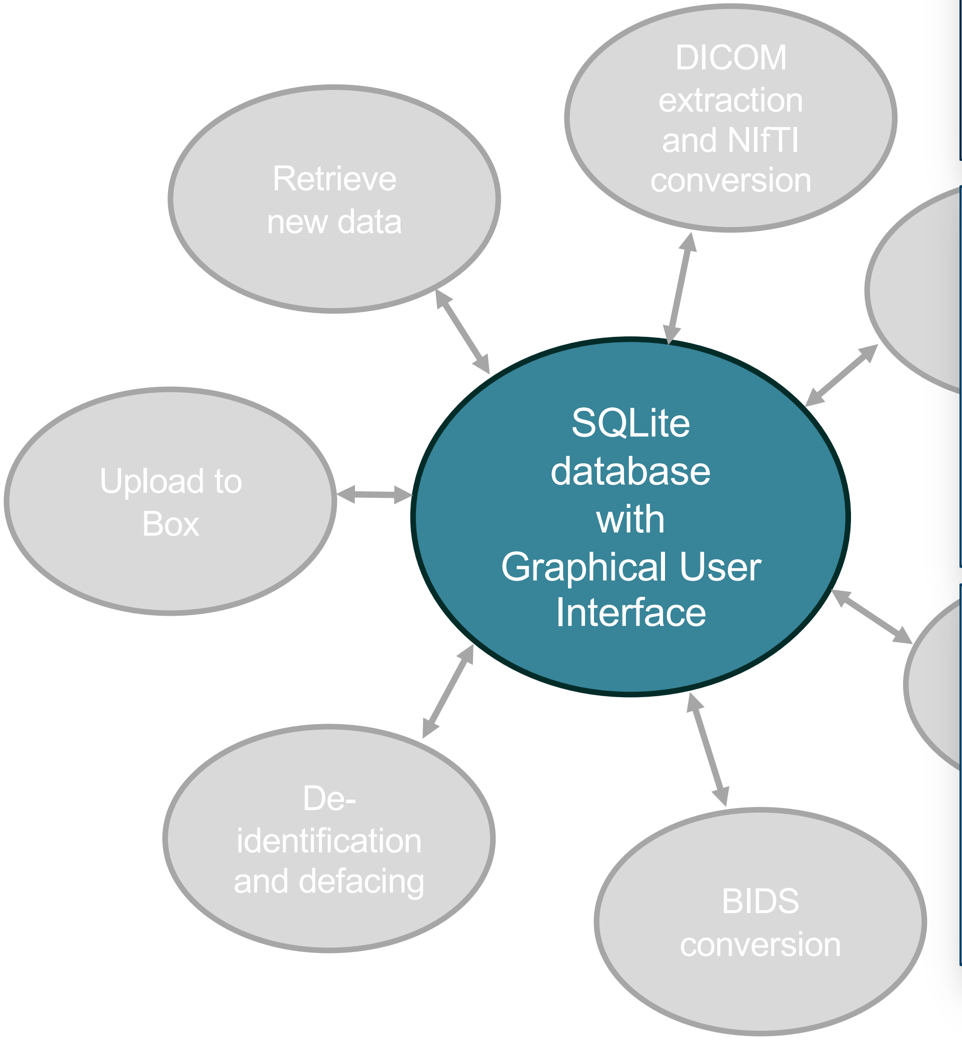
The software is freely available on GitHub:
https://github.com/finetto/musc-automated-imaging-pipeline

Automated MRI Data Preprocessing Pipeline



- Can be run manually or on a schedule
- Modular design (steps can be enabled or disabled)
- Fully local (except data download and upload)
- Database stores information on each participant, session and series
- Database stores current processing state for each session

Database and GUI



SQLite database with Graphical User Interface

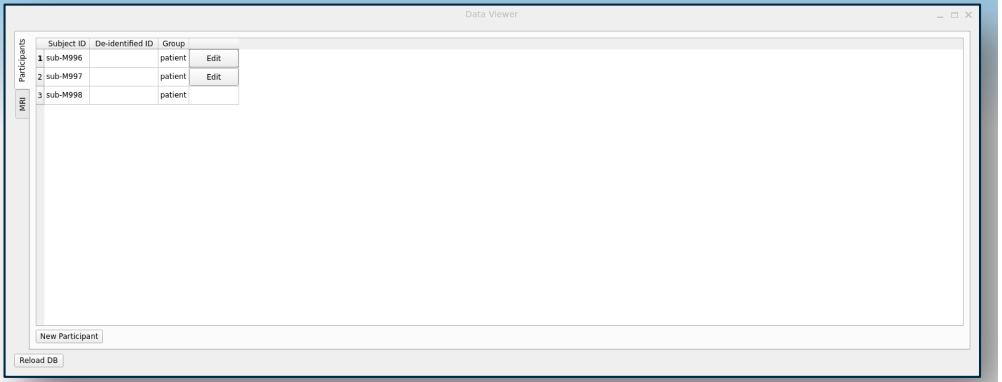
Retrieve new data

DICOM extraction and NIfTI conversion

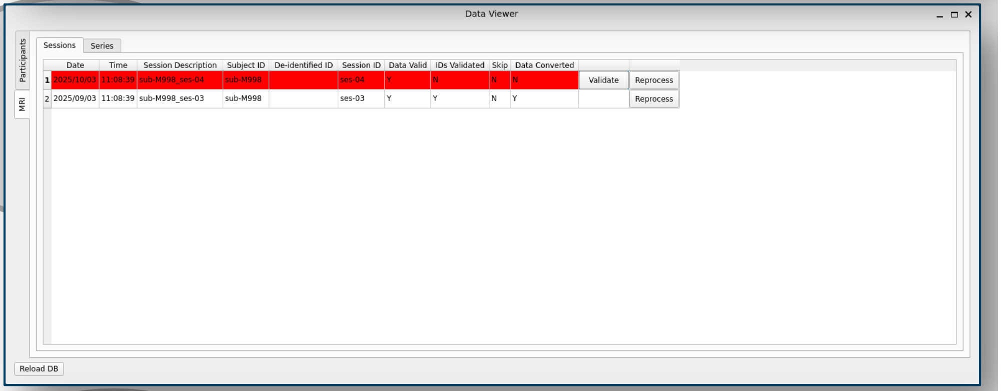
Upload to Box

De-identification and defacing

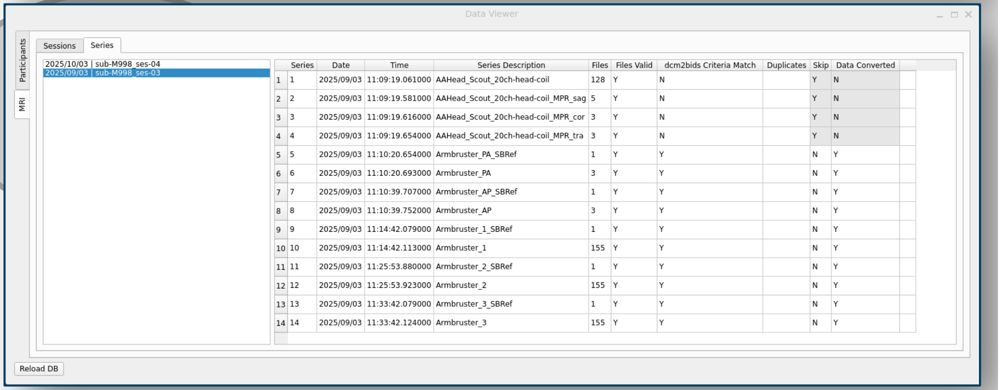
BIDS conversion



	Subject ID	De-identified ID	Group
1	sub-M996		patient
2	sub-M997		patient
3	sub-M998		patient



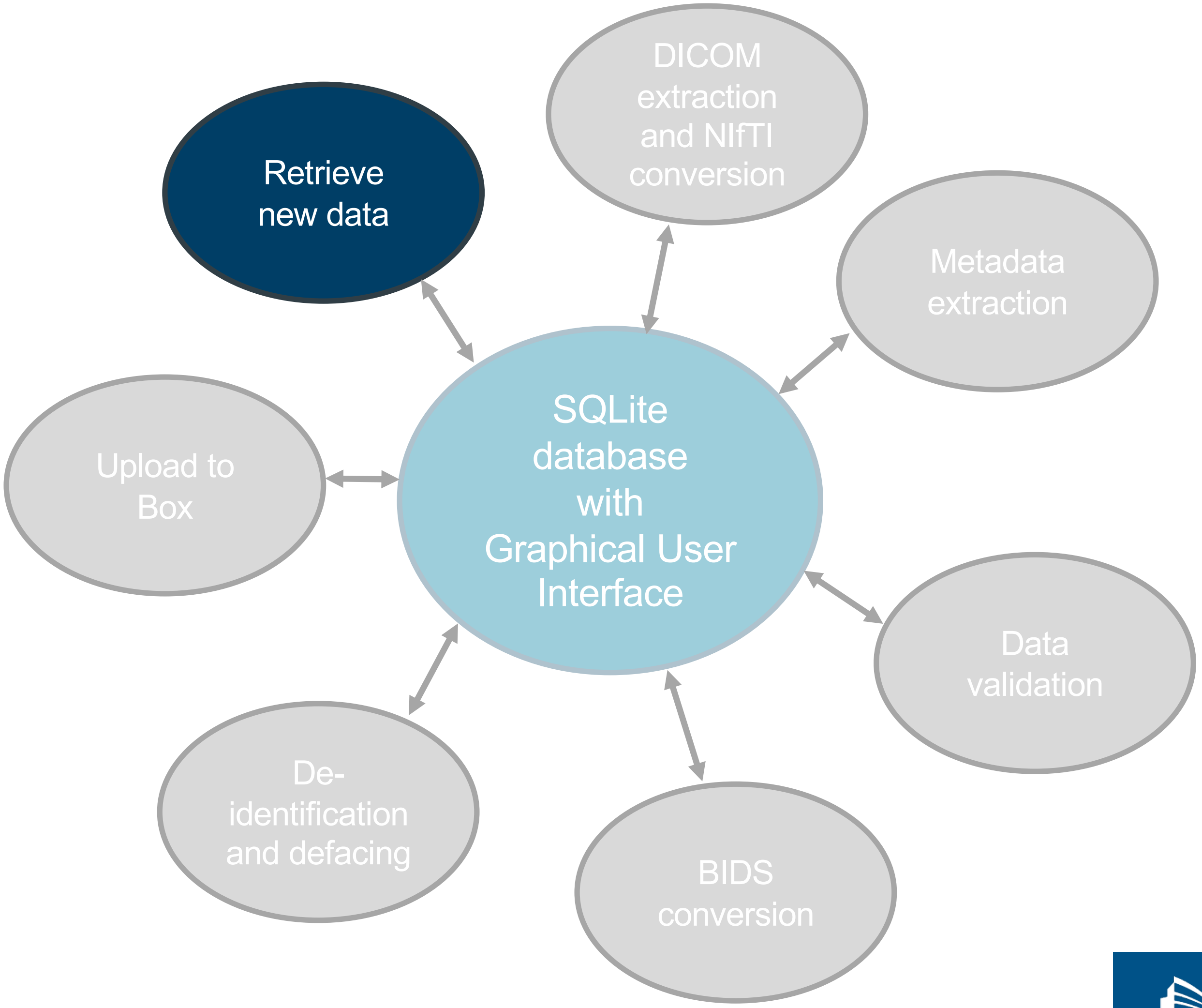
	Date	Time	Session Description	Subject ID	De-identified ID	Session ID	Data Valid	IDs Validated	Skip	Data Converted
1	2025/10/03	11:08:39	sub-M998_ses-04	sub-M998		ses-04	Y	N	N	N
2	2025/09/03	11:08:39	sub-M998_ses-03	sub-M998		ses-03	Y	Y	N	Y



2025/10/03 | sub-M998_ses-04
2025/09/03 | sub-M998_ses-03

	Series	Date	Time	Series Description	Files	Files Valid	dcm2bids Criteria Match	Duplicates	Skip	Data Converted
1	1	2025/09/03	11:09:19.061000	AAHead_Scout_20ch-head-coil	128	Y	N		Y	N
2	2	2025/09/03	11:09:19.581000	AAHead_Scout_20ch-head-coil_MPR_sag	5	Y	N		Y	N
3	3	2025/09/03	11:09:19.616000	AAHead_Scout_20ch-head-coil_MPR_cor	3	Y	N		Y	N
4	4	2025/09/03	11:09:19.654000	AAHead_Scout_20ch-head-coil_MPR_tra	3	Y	N		Y	N
5	5	2025/09/03	11:10:20.654000	Armbruster_PA_SBRef	1	Y	Y		N	Y
6	6	2025/09/03	11:10:20.693000	Armbruster_PA	3	Y	Y		N	Y
7	7	2025/09/03	11:10:39.707000	Armbruster_AP_SBRef	1	Y	Y		N	Y
8	8	2025/09/03	11:10:39.752000	Armbruster_AP	3	Y	Y		N	Y
9	9	2025/09/03	11:14:42.079000	Armbruster_1_SBRef	1	Y	Y		N	Y
10	10	2025/09/03	11:14:42.113000	Armbruster_1	155	Y	Y		N	Y
11	11	2025/09/03	11:25:53.880000	Armbruster_2_SBRef	1	Y	Y		N	Y
12	12	2025/09/03	11:25:53.923000	Armbruster_2	155	Y	Y		N	Y
13	13	2025/09/03	11:33:42.079000	Armbruster_3_SBRef	1	Y	Y		N	Y
14	14	2025/09/03	11:33:42.124000	Armbruster_3	155	Y	Y		N	Y

Data Retrieval



Retrieve new data

DICOM extraction and NIfTI conversion

Metadata extraction

Data validation

BIDS conversion

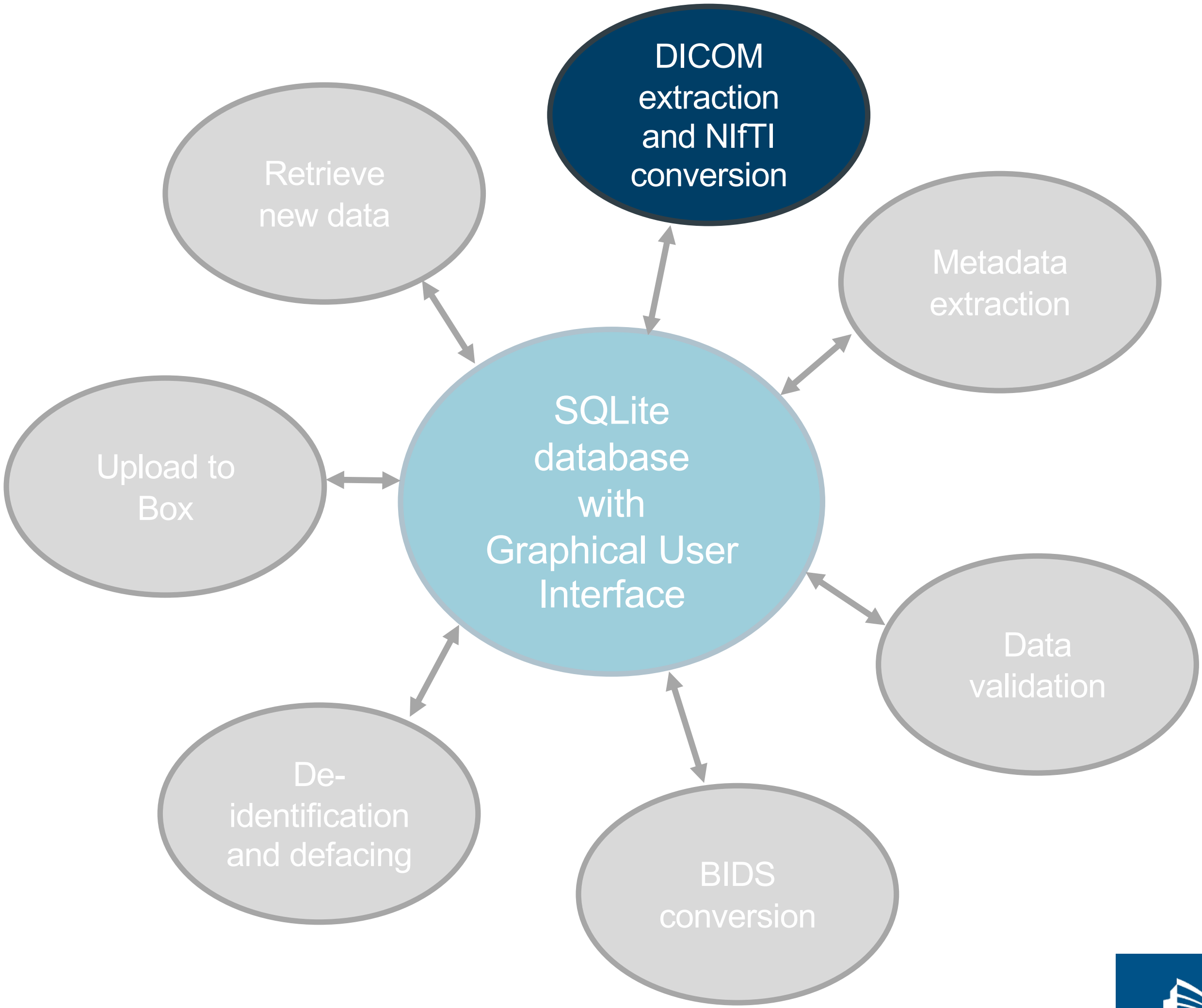
De-identification and defacing

Upload to Box

SQLite database with Graphical User Interface

- Scans the study folder on CBI home, or a local folder
- If a new session is found, the data is downloaded
- New sessions are added to the database, initiating all following steps

DICOM extraction and NIfTI conversion



DICOM extraction and NIfTI conversion

Retrieve new data

Metadata extraction

Upload to Box

SQLite database with Graphical User Interface

Data validation

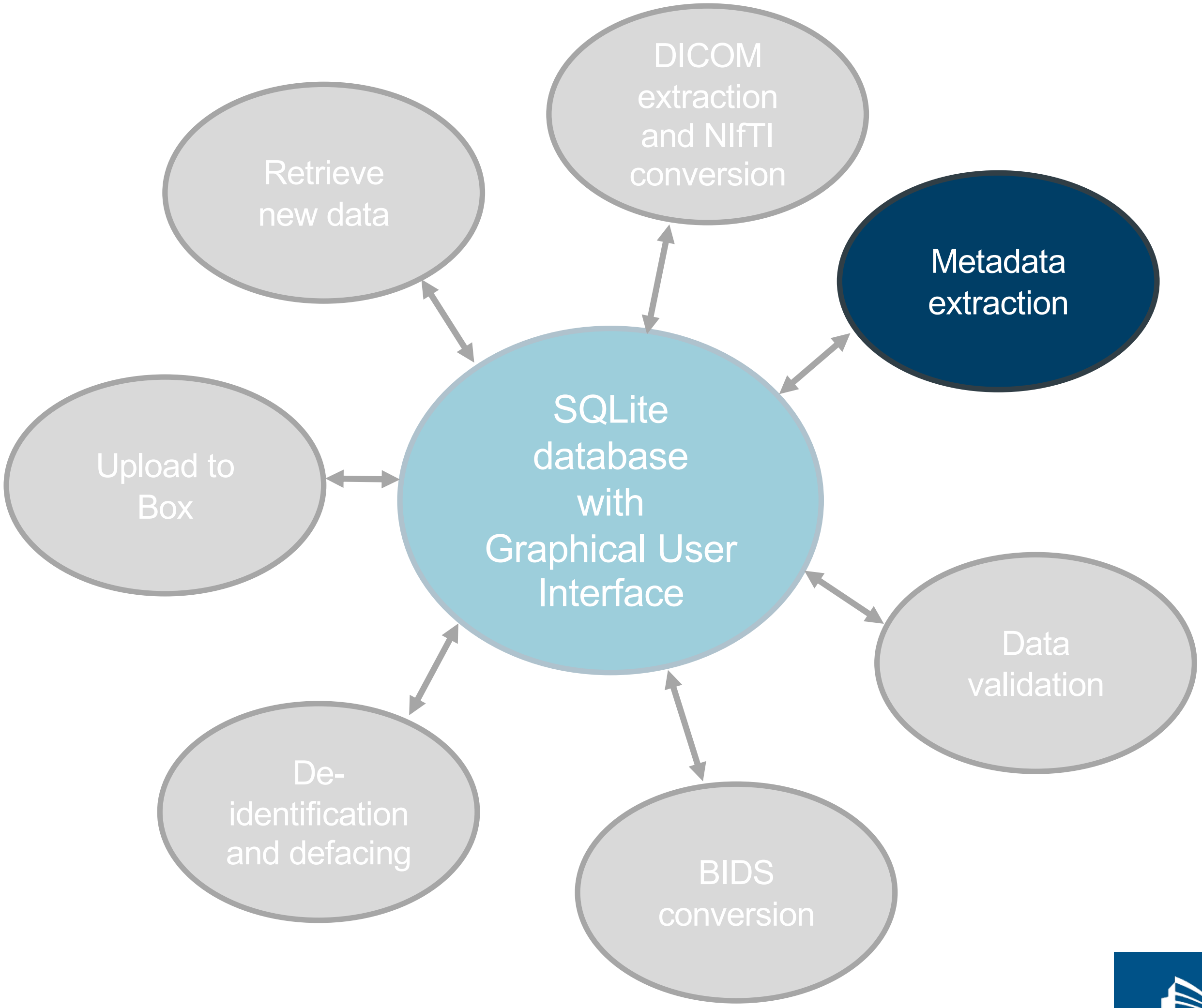
De-identification and defacing

BIDS conversion

- Extract the downloaded zip file
- Use dcm2niix [1] to convert to NIfTI
- Get all series for the session and add them to the database

[1] https://github.com/rordenlab/dcm2niix

Metadata extraction


Retrieve new data

DICOM extraction and NIfTI conversion

Metadata extraction

SQLite database with Graphical User Interface

Upload to Box

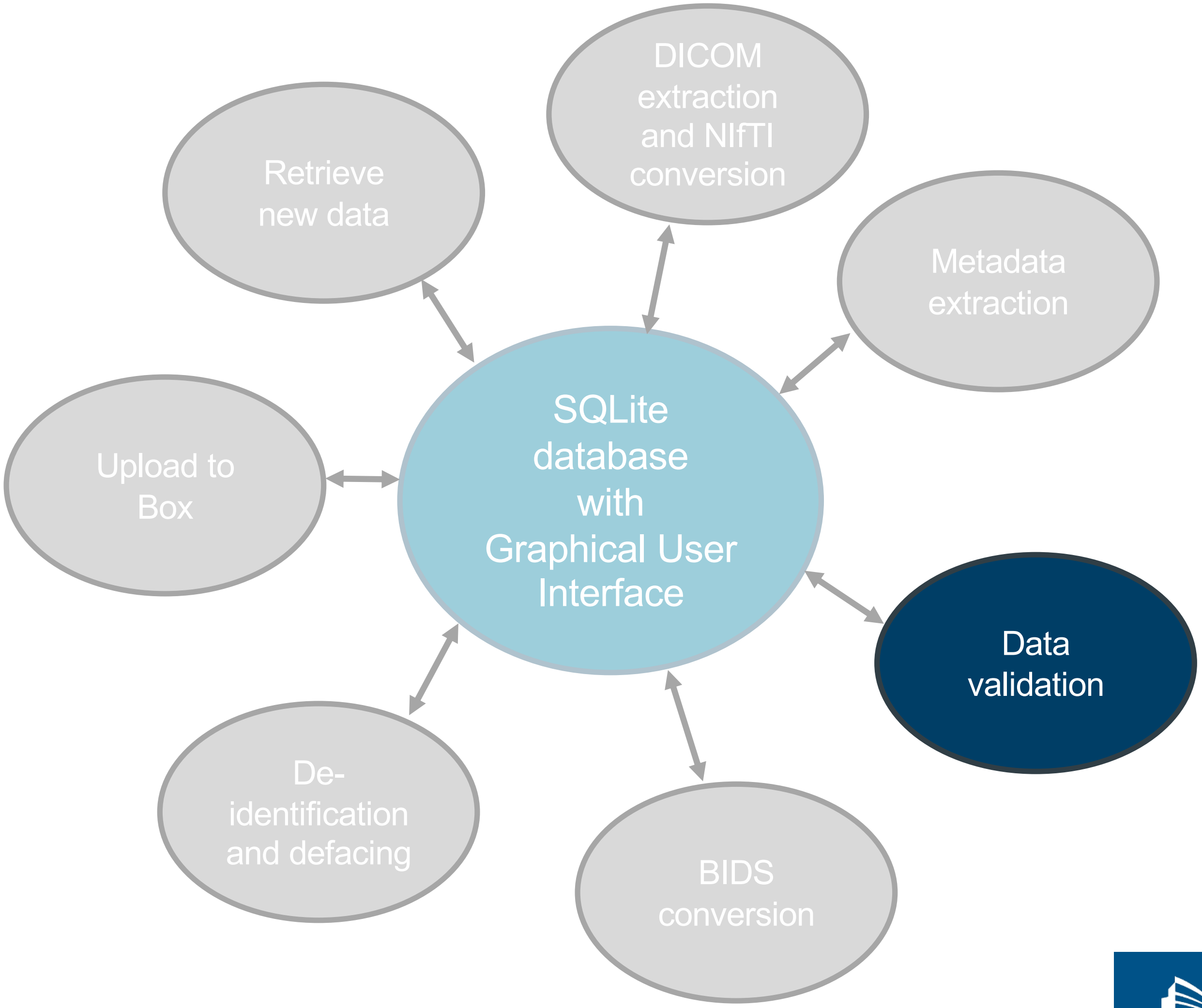
Data validation

De-identification and defacing

BIDS conversion

- Extract timestamp, subject ID and session ID from data file name
- Extract series information from sidecar file

Data validation


Retrieve new data

DICOM extraction and NIfTI conversion

Metadata extraction

Upload to Box

SQLite database with Graphical User Interface

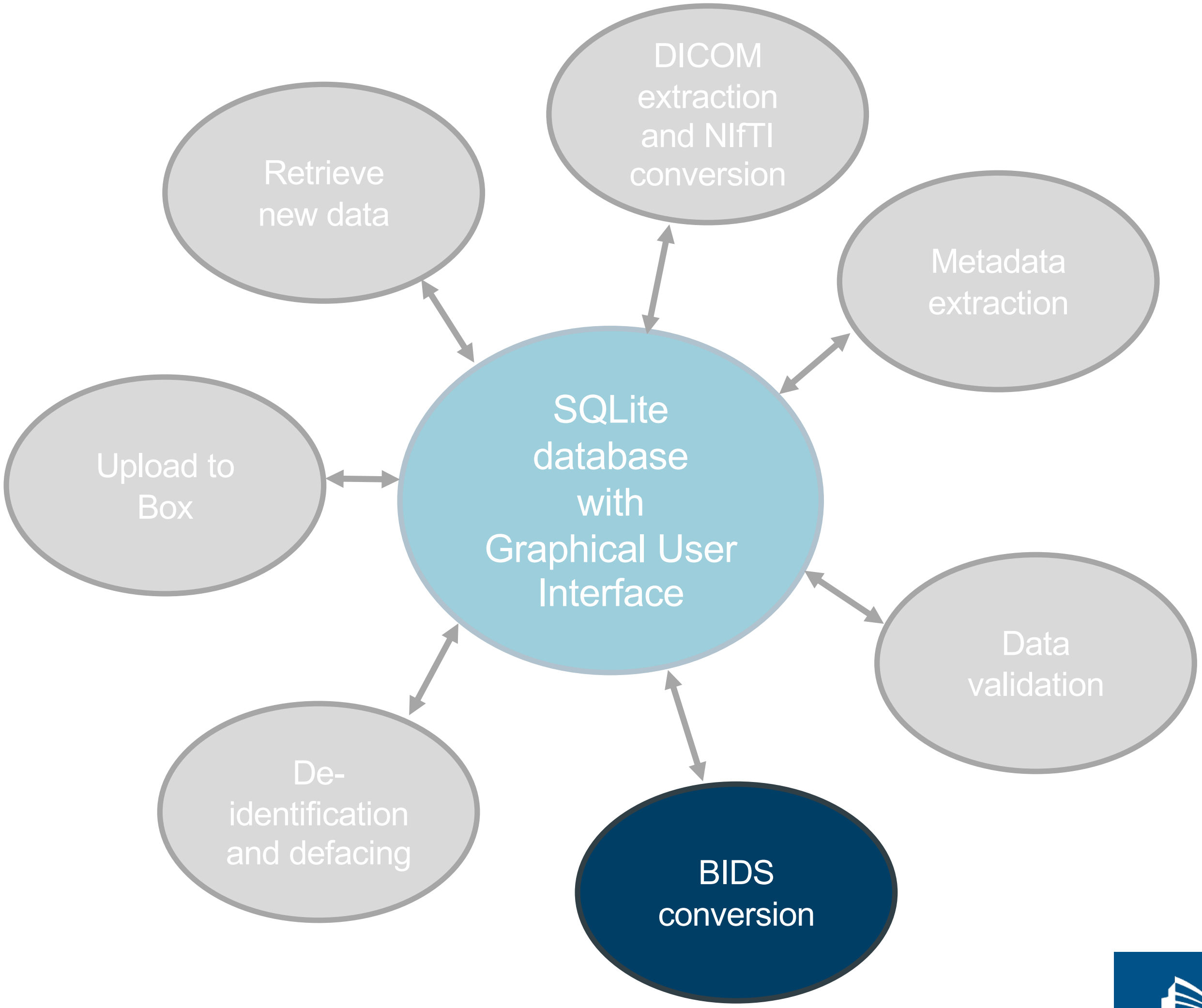
Data validation

De-identification and defacing

BIDS conversion

- Make sure files were extracted correctly
- Match dcm2bids criteria to each series
- Identify duplicate series (only one will be processed)
- Make sure the extracted series match the CBI Summary file (if available)
- Ask user to confirm subject ID and session ID
- User can mark sessions or series to skip

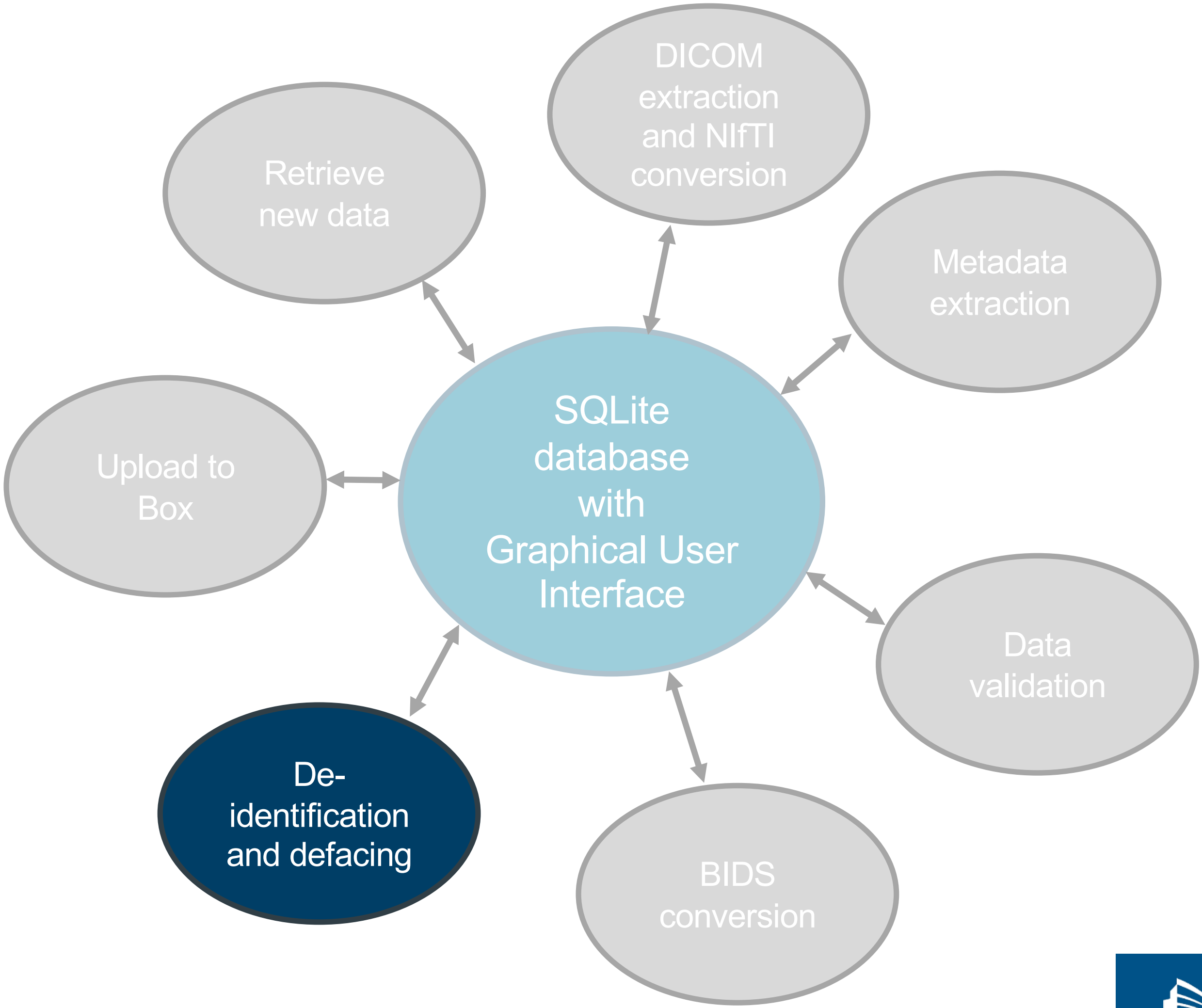
BIDS conversion



- Sessions that passed data validation will be converted to BIDS format
- NIfTI files are converted using dcm2bids [2]
- User needs to define a dcm2bids configuration file

[2] https://unfmontreal.github.io/Dcm2Bids/3.2.0/

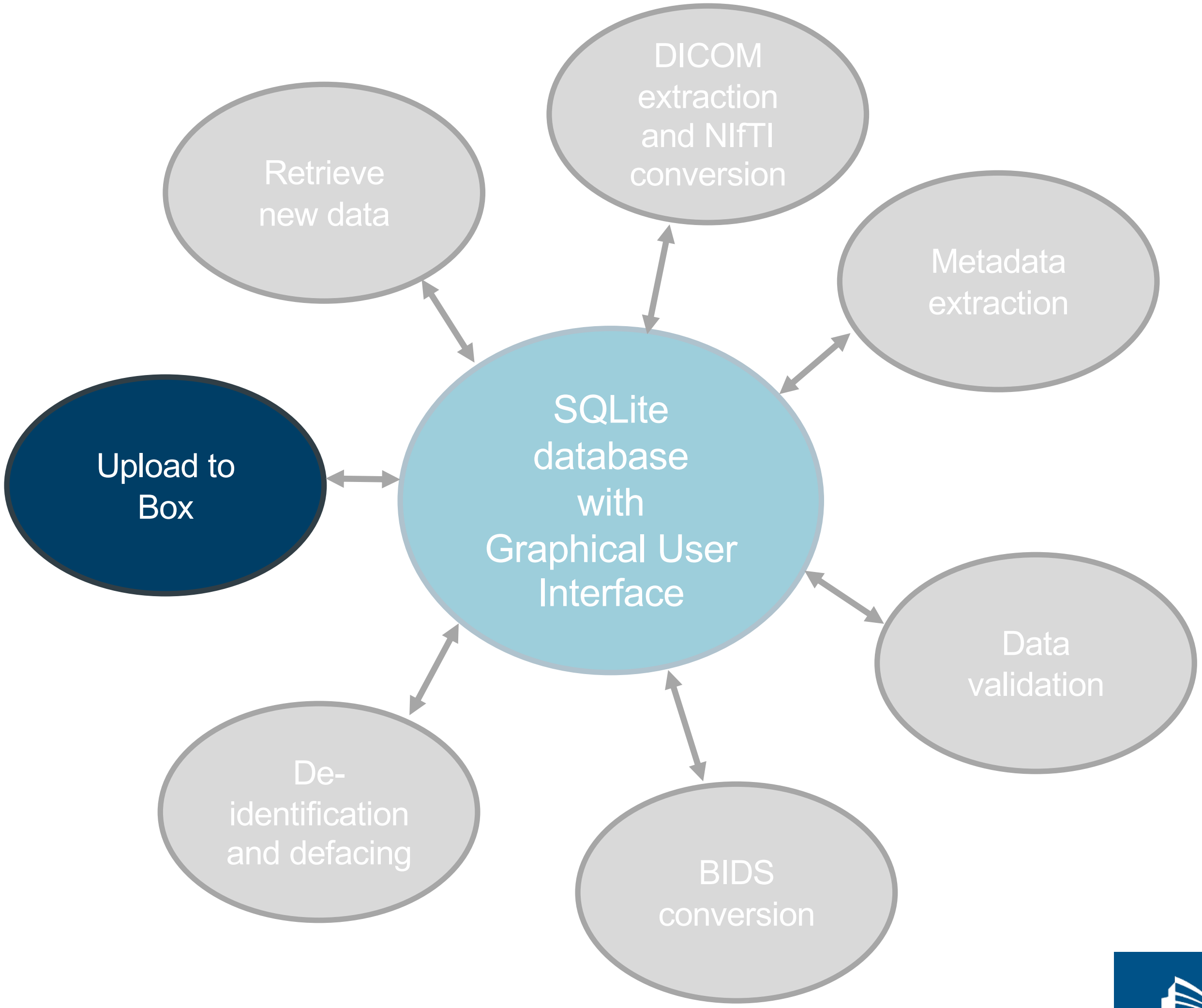
De-identification and defacing



- Optional step
- De-identified IDs are either assigned to subject or automatically generated
- All files are renamed using de-identified subject IDs
- Anatomical images are defaced using FSL [3] and pydeface [4]
- De-identified data is stored in separate folder structure

[3] https://fsl.fmrib.ox.ac.uk/fsl
[4] https://github.com/poldracklab/pydeface

Data upload to Box


SQLite database with Graphical User Interface

Retrieve new data

DICOM extraction and NIfTI conversion

Metadata extraction

Data validation

BIDS conversion

De-identification and defacing

Upload to Box

- Optional step
- Data is uploaded to Box using Box SDK [5]
- Source data, converted data and de-identified data are uploaded to separate folders
- For each file, the upload is verified and repeated if necessary
- Database and config files are backed up to Box as well

[5] https://github.com/box/box-python-sdk-gen

Installation and configuration

Installation

- Tested on Ubuntu and WSL2
- Instructions provided in the repository *README* file
- To install, download/clone the repository and run the *install.sh* script

Configuration

- There are several configuration files that need to be added and/or modified
  - Database configuration
  - Data download configuration (CBI Home or local folder)
  - Study configuration (e.g. study name, format of IDs,…)
  - Processing configuration (e.g. destination folders)
  - Box configuration
  - Notifications
- If synchronizing from CBI Home, one needs to create a credentials file
- A dcm2bids configuration file is necessary for the BIDS conversion

Running the pipeline

- The pipeline can be run on a schedule (e.g. every hour). This is configured automatically when running the installation script
- The pipeline can also be run manually with the *run_mri_pipeline.sh* script
- To launch the Data Viewer GUI, one can use the *run_data_viewer.sh* script

Software demo