

A System Architecture for Sharing De-Identified, Research-Ready Brain Scans and Health Information Across Clinical Imaging Centers



Ann L. Chervenak, Theo G.M. van Erp, Carl Kesselman,
Mike D'Arcy, Janet Sobell, David Keator, Lisa Dahm,
Jim Murry, Meng Law, Anton Hasso, Joe Ames,
Fabio Macciardi and Steven G. Potkin

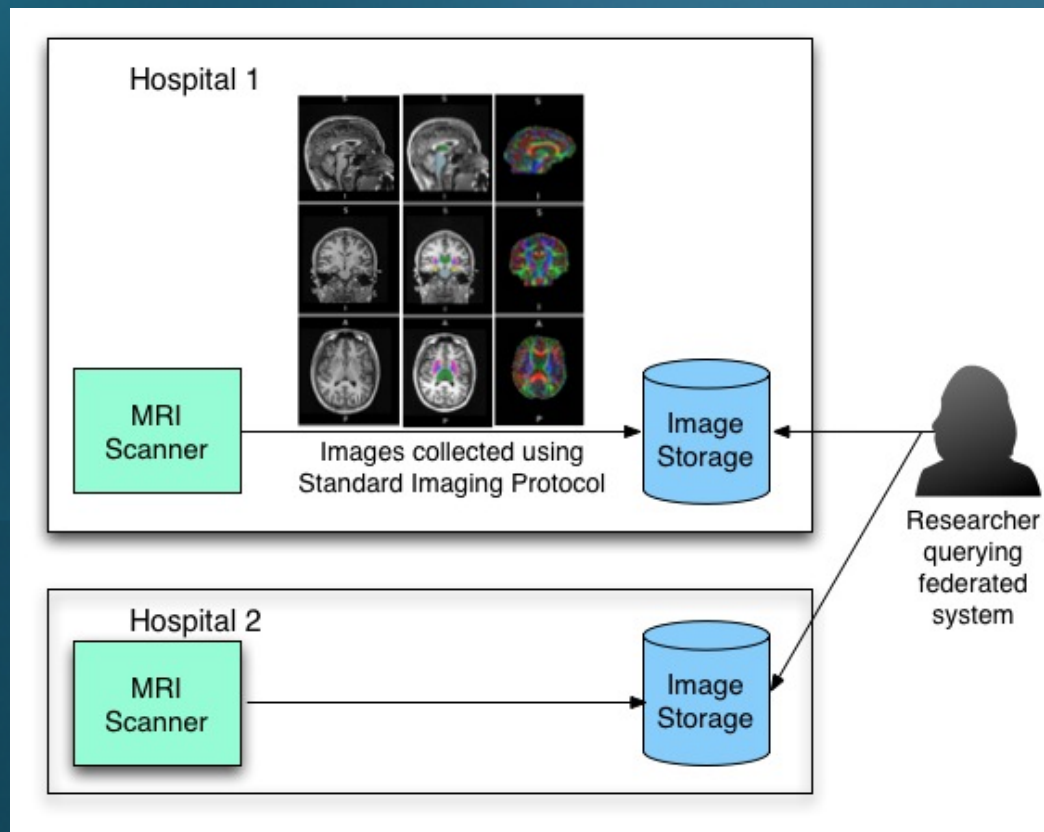
University of Southern California

University of California at Irvine

Challenge: Collecting Brain Images to Benefit both Research and Clinical Practice

- Research on complex brain disorders:
 - Increasingly relies on the collection of large standardized brain magnetic resonance imaging (MRI) data sets
 - Expensive for research studies
- Clinical MRI scanners in hospitals collect a large number of scans used for diagnosis of illness and injury:
 - On average, approximately 2,500 clinical brain scans per year
- Typically cannot use clinical scans for research purposes
 - Use of different scanners and imaging protocols at different facilities
 - Lack of optimization of scan protocols for research purposes

Goal: Develop a Research-Ready Brain Imaging Resource



- Acquire standardized, research-ready clinical brain scans
- Aggregate these scans across multiple imaging facilities
- Provide a convenient user interface for use by researchers and clinical radiologists

Potential Benefits of Collecting Research-Ready Clinical Brain Scans

- Benefits to research:
 - Access to large number of research-quality scans that are otherwise costly to obtain
 - Compare and contrast patients with similar conditions
 - Cohort identification
 - Example: Researcher reviews treatment outcomes for MS patients over time
 - Query for patients with MS based on their diagnosis codes
 - Patients scanned at first diagnosis and at 6 month intervals
 - Compare scans of patients who received different treatments
 - At no additional cost, can compare treatment outcomes

Benefits (continued)

- Benefits to clinical practice:
 - Radiologists reviewing scans can compare and contrast with scans from other patients with similar demographic, diagnostic, and treatment status
 - Compare quantitative measures (e.g., brain volumes) that are commonly compared in research studies
 - Example: Physician interpreting an MRI scan reviews other patient scans with similar characteristics

Contributions of Pilot Project

- (1) Defined a standardized brain imaging protocol
 - In use at the University of Southern California (USC) and the University of California at Irvine (UCI) hospitals
- (2) Developed general, open source, end-to-end software:
 - Securely store and manage imaging data at each institution
 - Support data federation across institutions, including distributed image query and download
 - User interface for radiologists, researchers
- (3) Deployed pilot federation across USC and UCI
 - Supports retrieval of images based on subject characteristics
- (4) Explore models for patient consent with medical ethics teams from the pilot institutions

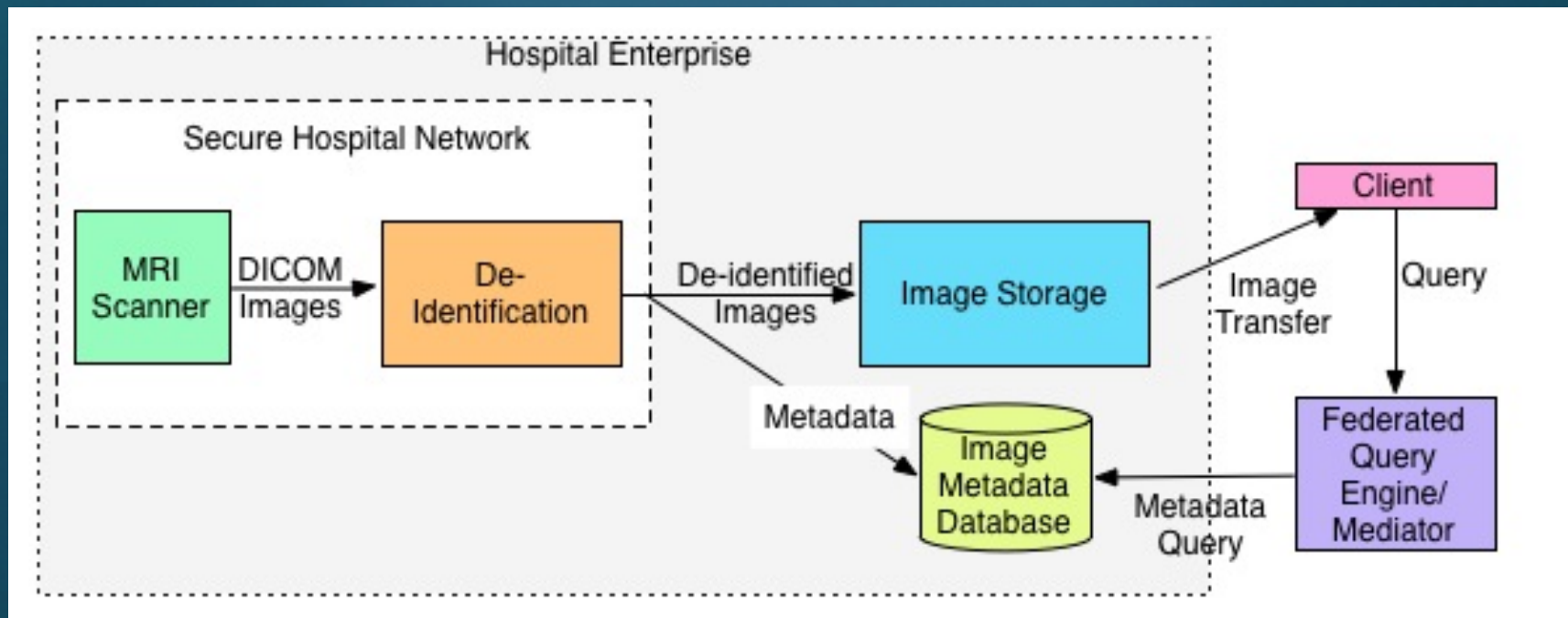
Defining a Standardized Brain Imaging Protocol

- Required expertise and cooperation of clinical neuroradiologists at both hospitals
 - Their primary responsibility is accurate patient diagnosis
- Four factors in protocol design:
 1. Minimal time impact on typical clinical brain MRI acquisitions (5 to 10 minutes)
 2. Portability so sequences can be implemented across most widely used scanner platforms (GE, Phillips, Siemens)
 3. Acquisition of whole brain scans
 4. Use of high quality sequences that will enable accurate, semi-automated image analyses

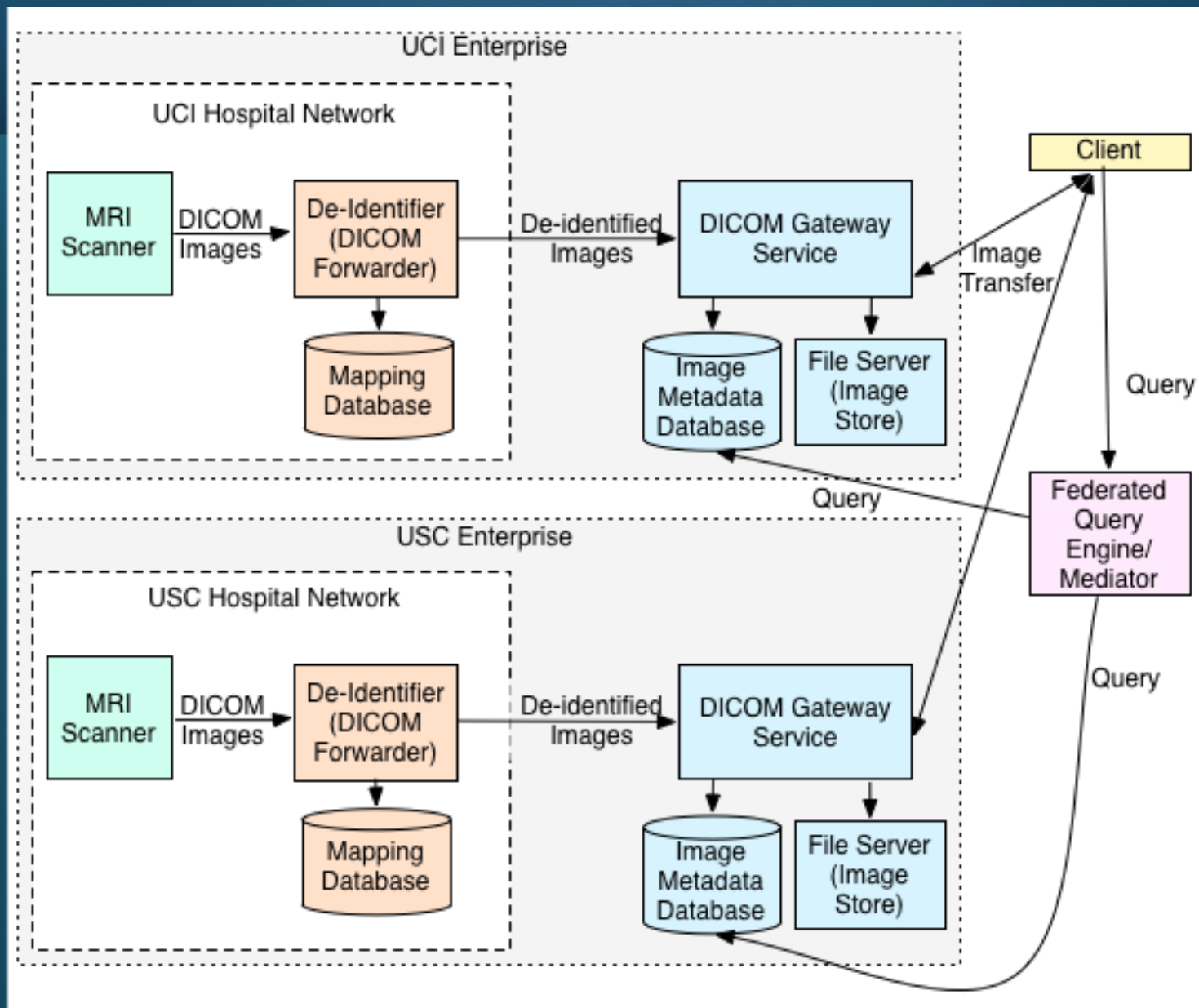
Standardized Imaging Protocol Sequence

- Collected on all clinical brain scans (18+ yrs, non-emergency) where feasible
- Diffusion Tensor Imaging (DTI):
 - Allows for study of white matter anatomy
- High-resolution T1:
 - Allows for quantification of gray matter, white matter, and cerebral spinal fluid volumes
- B1:
 - Represents homogeneity of magnetic field in scanned areas
 - Intensities determined by tissue type
- MagPhan Phantom Sequence
 - Used for correcting spatial distortion and inhomogeneity in images
 - Collected weekly at each site

High-Level View of Image Capture and Query



Architecture Overview

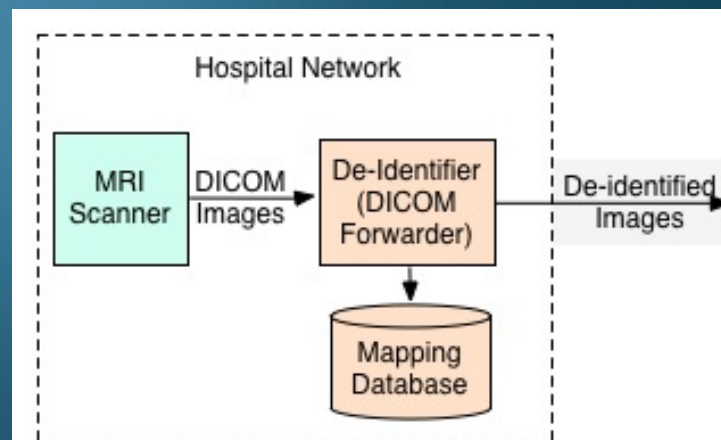


Three Main Functional Components

- DICOM (Digital Imaging and Communications in Medicine) Forwarder:
 - Runs on secure hospital network
 - De-identifies images coming from the scanner
 - Forwards them for storage and access
- DICOM Image Gateway:
 - Runs in hospital enterprise
 - Stores de-identified images
 - Extracts metadata into a searchable database
 - Responds to metadata queries and serves images
- Federated Query Engine/Mediator:
 - Distributes user queries for images with certain metadata attributes to the sites in the federation

DICOM Forwarder

- Interface between the scanner and federated imaging repository
- Receives DICOM image files pushed from the local scanner
- Removes patient identifying information in accordance with:
 - HIPAA (*Health Insurance Portability and Accountability Act*) guidelines for human subjects research
 - Institutional Review Board (IRB) restrictions for this project
- Forwards the images to the Gateway service for storage and discovery



De-Identification of DICOM Image Files

- Removes non-HIPAA-compliant fields
- Replaces other fields with “anonymous” tokens
- Remaps identifier fields with newly-minted unique IDs
- Forwarder stores mappings between original fields and remapped attributes in a local mapping database
 - Possible for authorized staff in hospital network to do a reverse mapping if necessary
- Researchers and clinicians using the federated service do not have access to any Personal Health Information

Linking De-Identified Images with Electronic Medical Records Data

- To support richer queries on images, IRB allows us to:
 - Use the medical record identifiers and scan dates stored in the secure mapping table
 - Query for IRB-approved health information associated with the scan from the Electronic Medical Records (EMR) systems
- We are allowed to obtain:
 - Diagnosis (ICD-9 codes)
 - Age
 - Sex
 - Race
 - Ethnicity

Architecture Overview

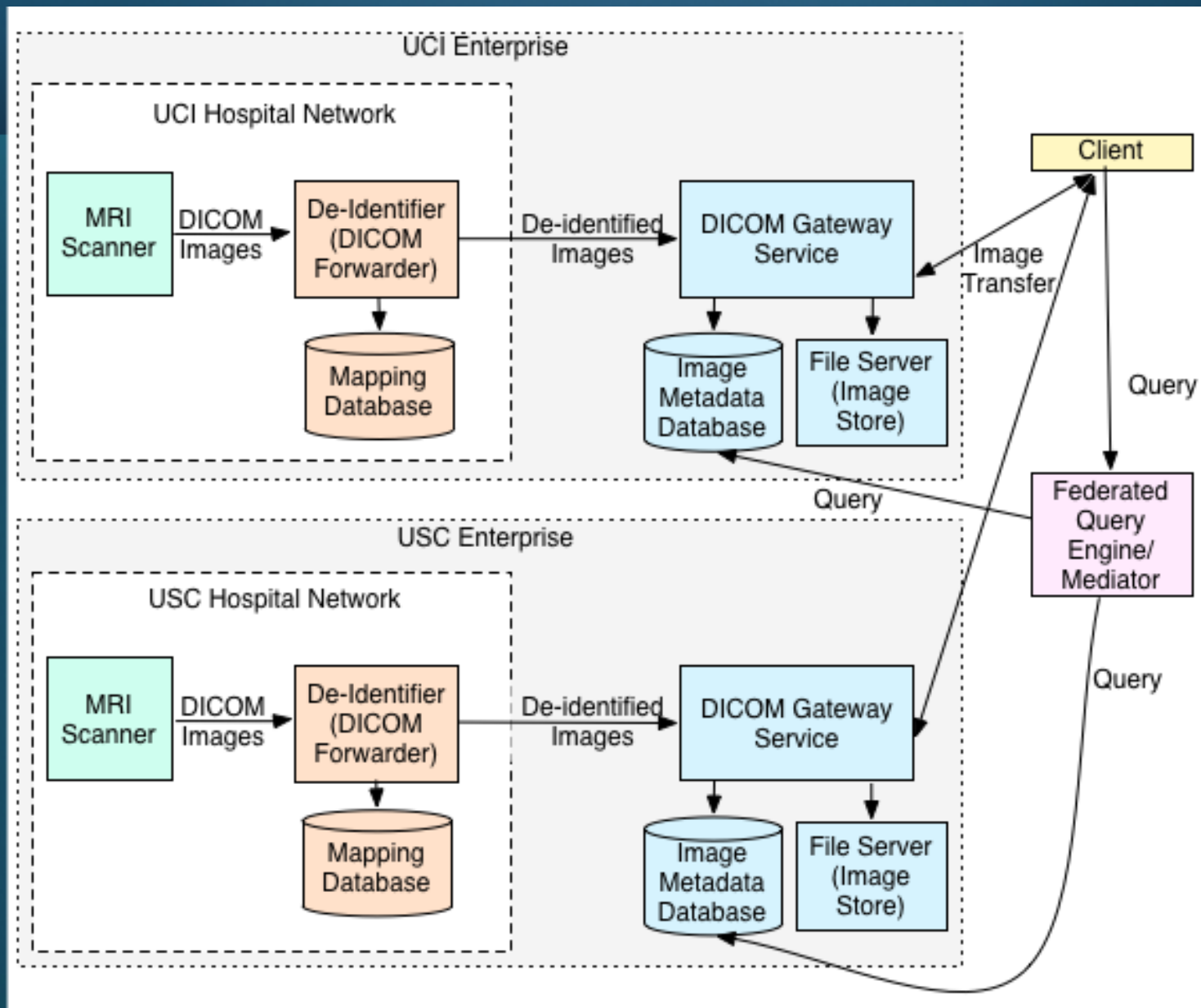
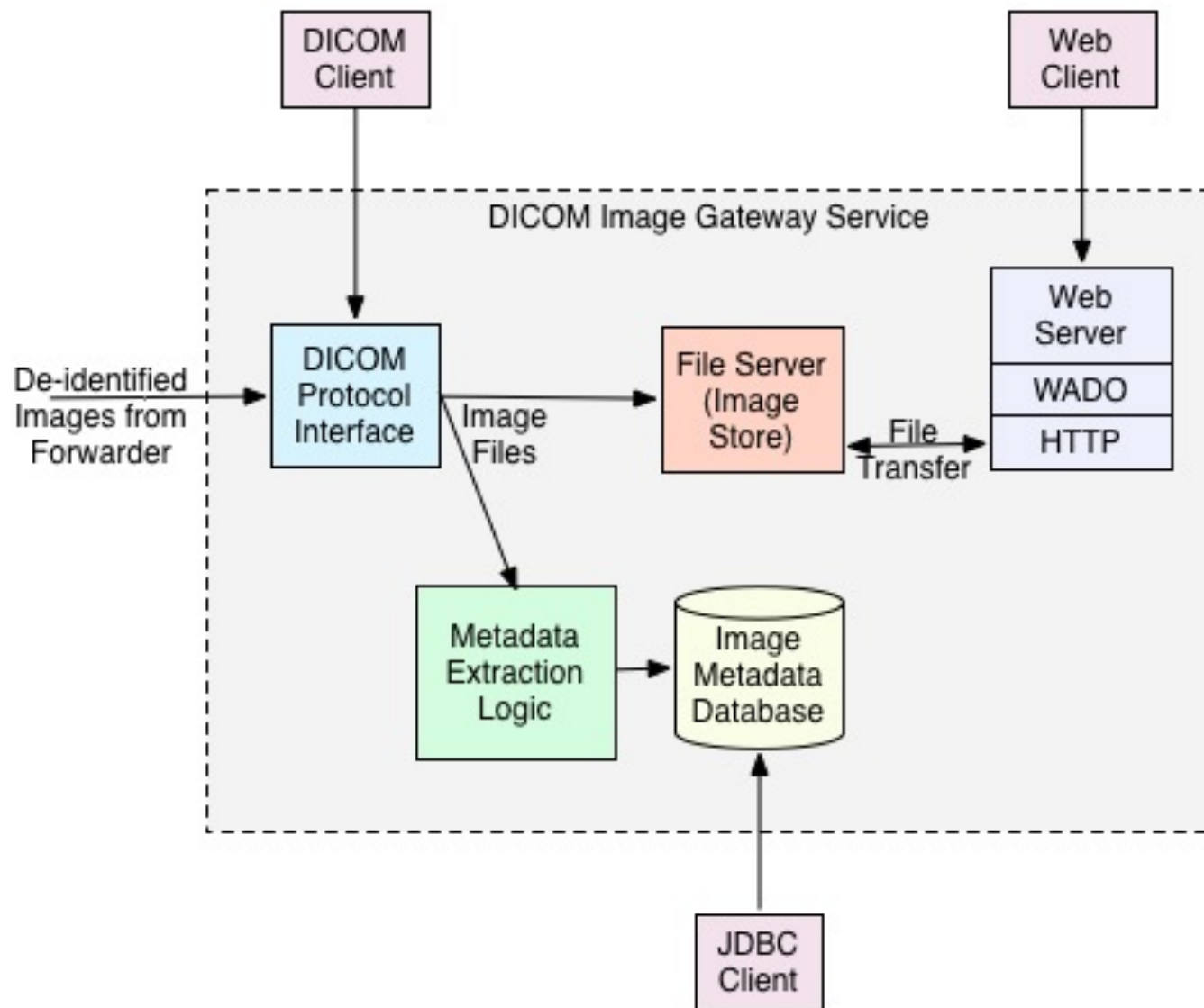


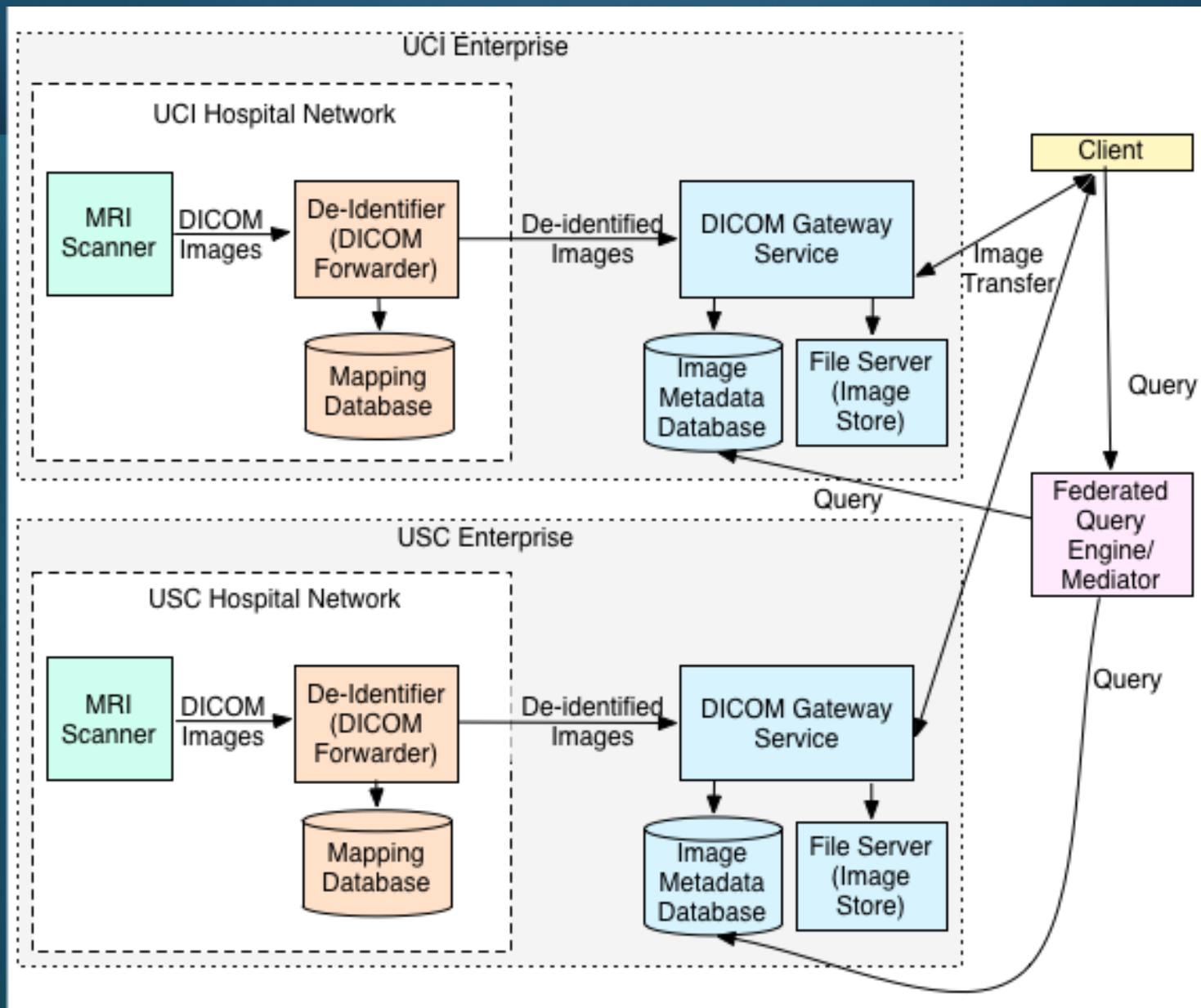
Image Gateway Service



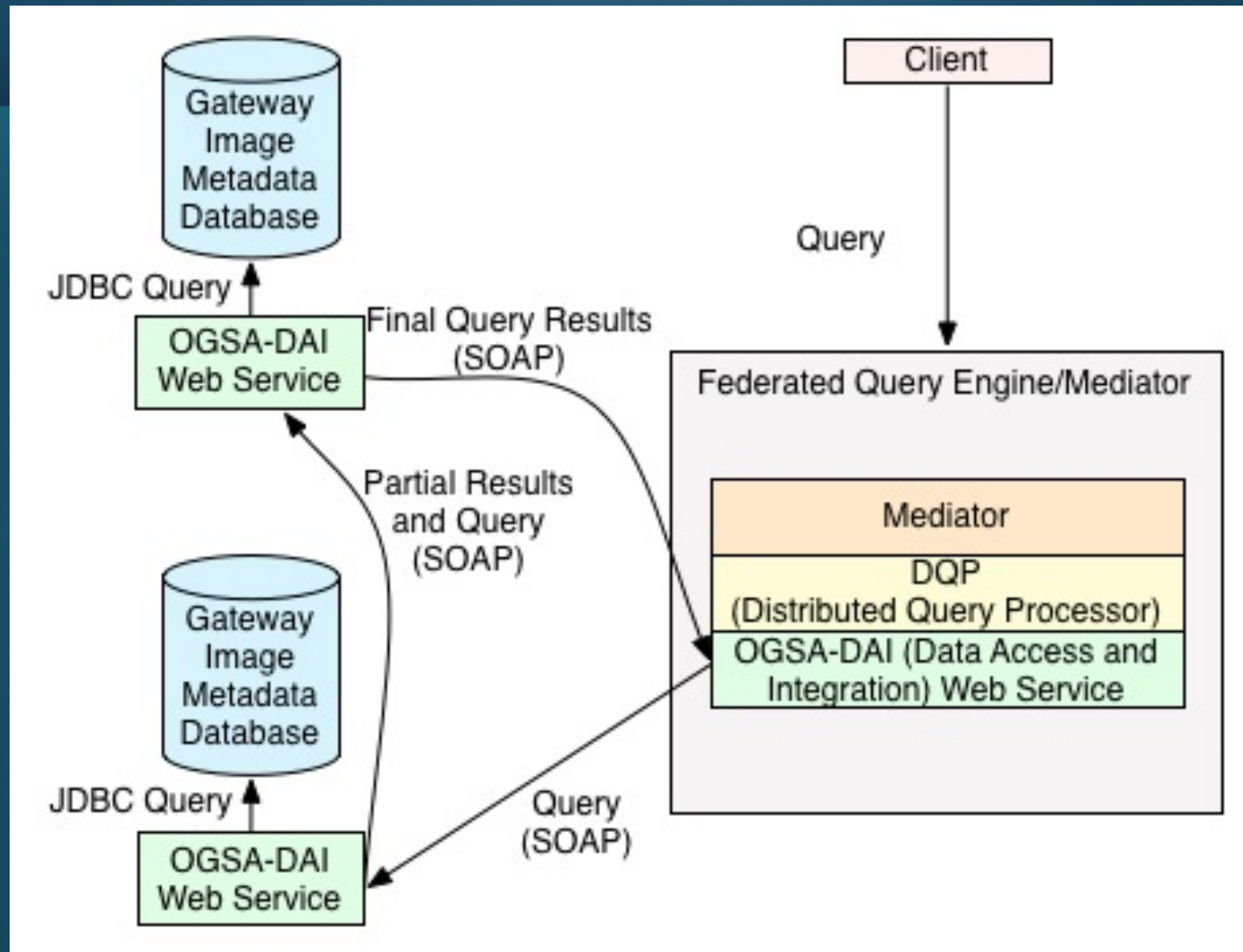
Four Components in Gateway Service

1. DICOM protocol interface: allows clients to interact with gateway service as with any DICOM service, such as a PACS (Picture Archiving and Communication System) system
2. Stores complete file images in a file server
 - They can be accessed by file transfer protocols
3. Metadata storage
 - Extracts metadata from DICOM image headers
 - Stores in a database, responds to queries on metadata attributes
4. Web server that provides alternate means of query and retrieval for images

Architecture Overview



Federated Query Engine/Mediator



Federated Query Engine/Mediator

- Accepts client queries for images with specified metadata attributes
- Distributes those queries across the institutions in the federation
 - Submits queries to Image Metadata Database in each DICOM Gateway Service
- Components:
 - OGSA Data Access and Integration Service (OGSA-DAI) provides secure access to Gateway's Metadata Database
 - OGSA Distributed Query Processing Service (DQP) federates DAI web service interfaces
 - Mediator: builds a global, virtual database view of the metadata databases in the federation

Query Interface

CTSI SIP DICOM Q/R

CTSI SIP DICOM Q/R

https://ctsi-sip-2.isi.edu:8443/dcmweb/

Google

Image Browser

Data Browser

Query

Match All

Body part

equals (disregard case)

head

Search

Clear Results

Clear All

Export

Export Type: CSV (Excel)

☒ All Fields

☐ Show in Window

Source Database ^	Patient ID	Age	Weight	Sex	Series number	Modality	Body part	Series description
▶ UCI CTSI SIP	4935fe76-7387-11e1-bf5d-00 058Y		77.110712709	F	7	MR	HEAD	AX T1 3T
▶ UCI CTSI SIP	4935fe76-7387-11e1-bf5d-00 058Y		77.110712709	F	9	MR	HEAD	COR GRE 3T
▶ UCI CTSI SIP	4935fe76-7387-11e1-bf5d-00 058Y		77.110712709	F	10	MR	HEAD	Mag_Images
▶ UCI CTSI SIP	4935fe76-7387-11e1-bf5d-00 058Y		77.110712709	F	11	MR	HEAD	Pha_Images
▲ UCI CTSI SIP	4935fe76-7387-11e1-bf5d-00 058Y		77.110712709	F	12	MR	HEAD	mIP_Images(SW)
Patient Name: ANONYMOUS								
Race: White								
Ethnicity: Hispanic								
Accession number: 01a9c9f3-8d75-11e1-8dfc-0050569200a4								
Study comments: PHANTOM_STUDY_UID=1.3.6.1.4.1.5962.99.1.3562041683.241822591.1327507105448.30818.0								
Station name: MRC35044								
Manufacturer: SIEMENS								
Manufacturer's model name: TrioTim								
Device serial number: 35044								
Software versions: syngo MR B17								
Institution name: UCI Medical Center 3T								
Institutional department name: Department								
Study Instance UID: 1.3.6.1.4.1.5962.99.1.2833971826.195908244.1334273866142.32259.0								
Series Instance UID: 1.3.6.1.4.1.5962.99.1.2833971826.195908244.1334273866142.32863.0								
Frame of reference UID: 1.3.6.1.4.1.5962.99.1.2833971826.195908244.1334273866142.32261.0								
ICD9 Code #1: 202.81								
ICD9 Description #1: OTHER MALIG LYMPHOMAS LYMPH NODES HEAD FACE&NECK								
ICD9 Code #2:								
ICD9 Description #2:								

Image Browser

CTSI SIP DICOM Q/R

CTSI SIP DICOM Q/R

https://ctsi-sip-2.isi.edu:8443/dcmweb/

Most Visited Getting Started Latest Headlines Google Maps YouTube News Wikipedia Bookmarks

Image Browser Data Browser Query

Navigation

- CTSI_B1_Body DO NOT CHANGE
- CTSI_B1_Head DO NOT CHANGE
- CTSI_DTI DO NOT CHANGE_TRACEW
- AX DIFFUSION 3T
- PhoenixZIPReport
- CTSI_DTI DO NOT CHANGE_ADC
- AX PD 3T
- COR GRE 3T
- POST AX T1 FAT MTC
- Pha_Images
- Pha_Images(SWI)

Total Patients: 352
Total Studies: 368
Total Series: 6978

Details

Series number:	7
Series description:	CTSI_DTI DO NOT CHANGE_TRACEW
Modality:	MR
Body part:	HEAD
Institution name:	UCI Medical Center 3T
Institutional department name:	Department
Station name:	MRC35044
Manufacturer:	SIEMENS
Manufacturer's model name:	TrioTim

Settings

Sort By: Image Number Sort Ascending: ☒

Image Scale: 2X Image Annotations: ☒

Image # 1
Series #: 7
Modality: MR
Dimensions: 96x96
Accession #: c2d9e18a-47e4-11e1-bf5d-0050569200a4

Image # 2
Series #: 7
Modality: MR
Dimensions: 96x96
Accession #: c2d9e18a-47e4-11e1-bf5d-0050569200a4

Image # 3
Series #: 7
Modality: MR
Dimensions: 96x96
Accession #: c2d9e18a-47e4-11e1-bf5d-0050569200a4

Image # 4
Series #: 7
Modality: MR

Image # 5
Series #: 7
Modality: MR

Image # 6
Series #: 7
Modality: MR

Related Work

- @neurlST project
 - Architecture supports federated data sets stored at multiple institutions by anonymizing and storing data at each site
 - Differs in standardized imaging protocol
- Extensible Neuroimaging Archive Toolkit (XNAT)
 - Open source imaging informatics
 - Projects can be federated into a single XNAT instance
 - Currently no built-in support for distributed queries across XNAT installations
- Medical Imaging Informatics Bench to Bedside (MI2B2)
 - Integrates DICOM imaging systems into I2B2 analytic data repository, used for cohort identification
 - Uses XNAT for image management, no federated query

Future Work

- Next stage will further integrate pilot system with clinical EMR systems: additional diagnostic metadata
 - Requires more extensive IRB approval, issues of informed consent
- Extending the SIP with additional scan sequences
 - Substitute for sequences already in use
- Neuroradiologists agree in principle on value of standardized, research-quality imaging protocol
 - In practice, may be reluctant to perform additional scans unless convinced of their clinical utility, no added malpractice liability
- Improve user interface based on radiologists' feedback
- Expand pilot federation to additional sites

Team

- UCI

- Theo G.M. van Erp
- David Keator
- Lisa Dahm
- Jim Murry
- Fabio Macciardi
- Steven G. Potkin
- Joseph Ames
- Anton Hasso
- Jason Handwerker
- Fred Greensite
- Dharmendra Patel

- USC: Information Sciences and Keck School of Medicine

- Mike D'Arcy
- Ann Chervenak
- Carl Kesselman
- Meng Law
- Janet Sobell
- Carlos Pato
- Michelle Pato