

# PARADIM: A digital infrastructure to support AI research in medical imaging

Yannick Lemaréchal,  
François Pelletier, Gabriel Couture, Pierre-Luc Asselin,  
Samuel Ouellet, Ronan Lefol, Luc Beaulieu  
and Philippe Després



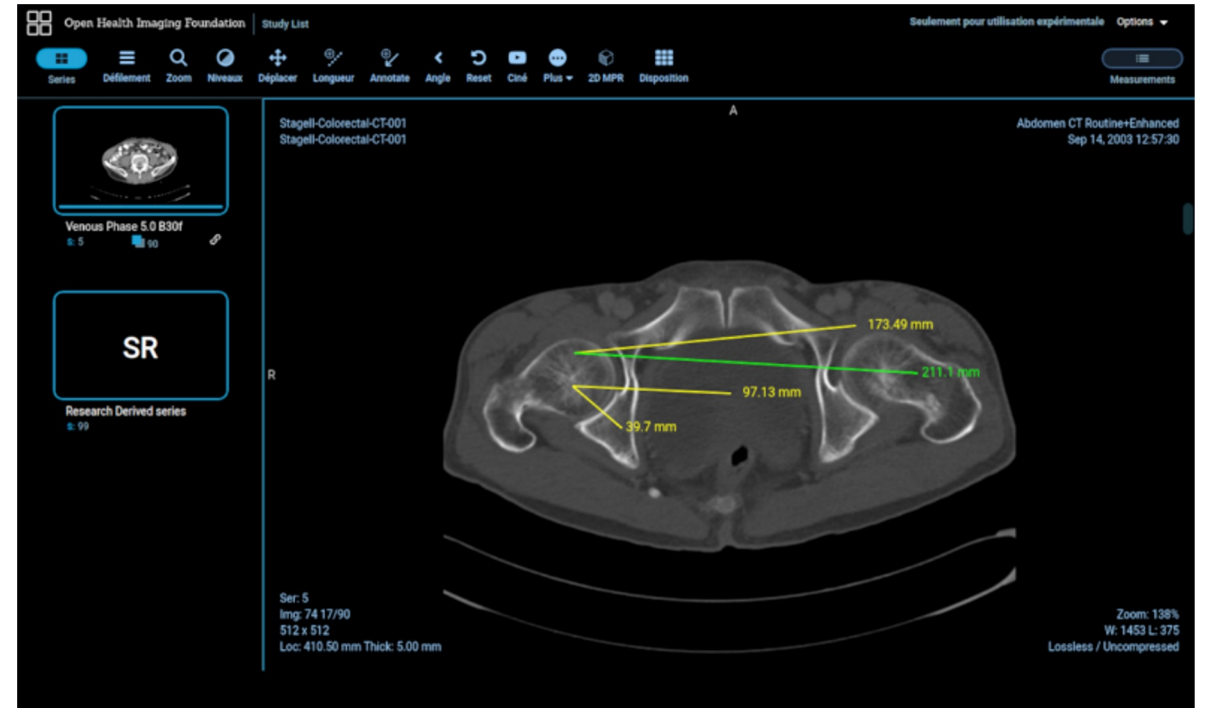
CENTRE DE RECHERCHE  
INSTITUT UNIVERSITAIRE  
DE CARDIOLOGIE  
ET DE PNEUMOLOGIE  
DE QUÉBEC  
UNIVERSITÉ LAVAL

---

# The need

Conducting AI research in medical imaging requires:

- complying with **complex ethical/legal framework**
- establishing ground truth for supervised machine learning (**capture expert annotations**)
- dealing with **large volumes of data**



# Context

npj | digital medicine

[Explore content](#) ▾ [About the journal](#) ▾ [Publish with us](#) ▾

---

[nature](#) > [npj digital medicine](#) > [perspectives](#) > article

[Perspective](#) | [Open Access](#) | [Published: 16 August 2019](#)

## The “inconvenient truth” about AI in healthcare

[Trishan Panch](#), [Heather Mattie](#) & [Leo Anthony Celi](#) 

[npj Digital Medicine](#) 2, Article number: 77 (2019) | [Cite this article](#)

53k Accesses | 75 Citations | 447 Altmetric | [Metrics](#)

[...] most healthcare organizations lack the **data infrastructure** required to collect the data needed to optimally train algorithms [...]

# Inspiration and guidance: Data-centric AI

Jun 16, 2021, 05:04pm EDT | 31,798 views

## Andrew Ng Launches A Campaign For Data-Centric AI



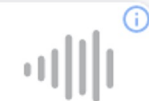
**Gil Press** Senior Contributor ⓘ

Enterprise & Cloud

*I write about technology, entrepreneurs and innovation.*



Listen to article 7 minutes



f

Data is eating the world so Andrew Ng wants to make sure we radically improve its quality.

t

“Data is food for AI,” says Ng, and he is launching a campaign to shift the focus of AI practitioners from model/algorithm development to the quality of the data they use to train

in

the models.

# Inspiration and guidance: MLOps

**MLOps** is a set of practices that aims to deploy and maintain machine learning models in production reliably and efficiently

**Objective: industrialize AI research in medical imaging**

# Inspiration and guidance : FAIR

## FAIR Principles

Sci. Data 3:160018 doi: 10.1038/sdata.2016.18 (2016)

### *Findable:*

F1 (meta)data are assigned a globally unique and persistent identifier;

**F2 data are described with rich metadata;**

F3 metadata clearly and explicitly include the identifier of the data it describes;

F4 (meta)data are registered or indexed in a searchable resource;

### *Interoperable:*

I1 (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2 (meta)data use vocabularies that follow FAIR principles;

I3 (meta)data include qualified references to other (meta)data;

### *Accessible:*

A1 (meta)data are retrievable by their identifier using a standardized communications protocol;

A1.1 the protocol is open, free, and universally implementable;

A1.2 the protocol allows for an authentication and authorization procedure, where necessary;

A2 metadata are accessible, even when the data are no longer available;

### *Reusable:*

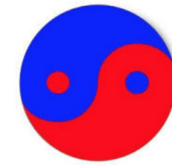
R1 meta(data) are richly described with a plurality of accurate and relevant attributes;

**R1.1 (meta)data are released with a clear and accessible data usage license;**

**R1.2 (meta)data are associated with detailed provenance;**

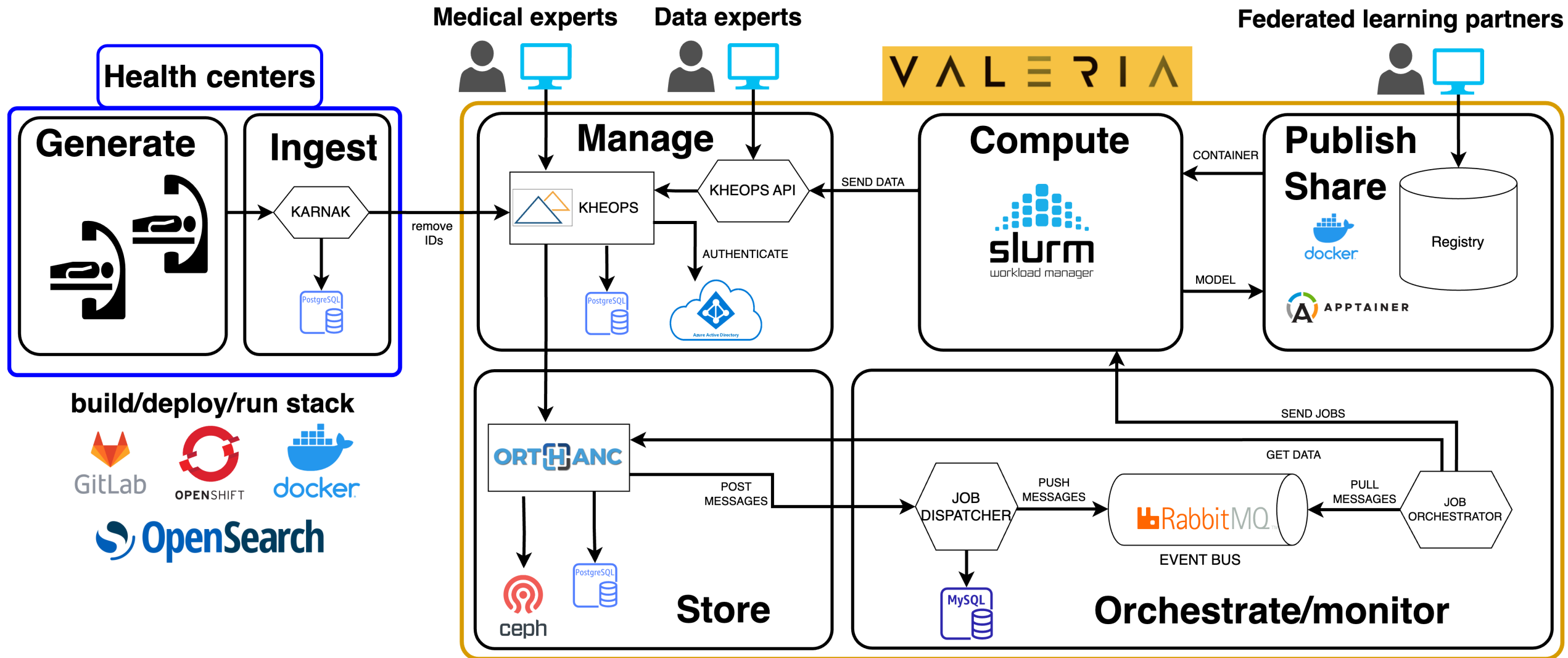
**R1.3 (meta)data meet domain-relevant community standards;**

Technology



**Domain-relevant  
content**

# The solution



# Ethical / legal challenges

## Challenges

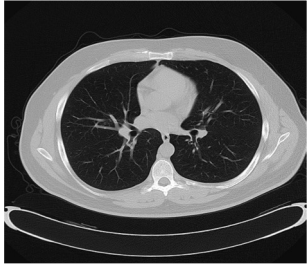
- ERB requires data destruction at the end of project
- Logging data accesses is now mandatory (Québec law)
- De-identifying images is not trivial (hundreds of fields in DICOM files)
- Current model involves sending data to third parties (and losing control over it)

## Solution

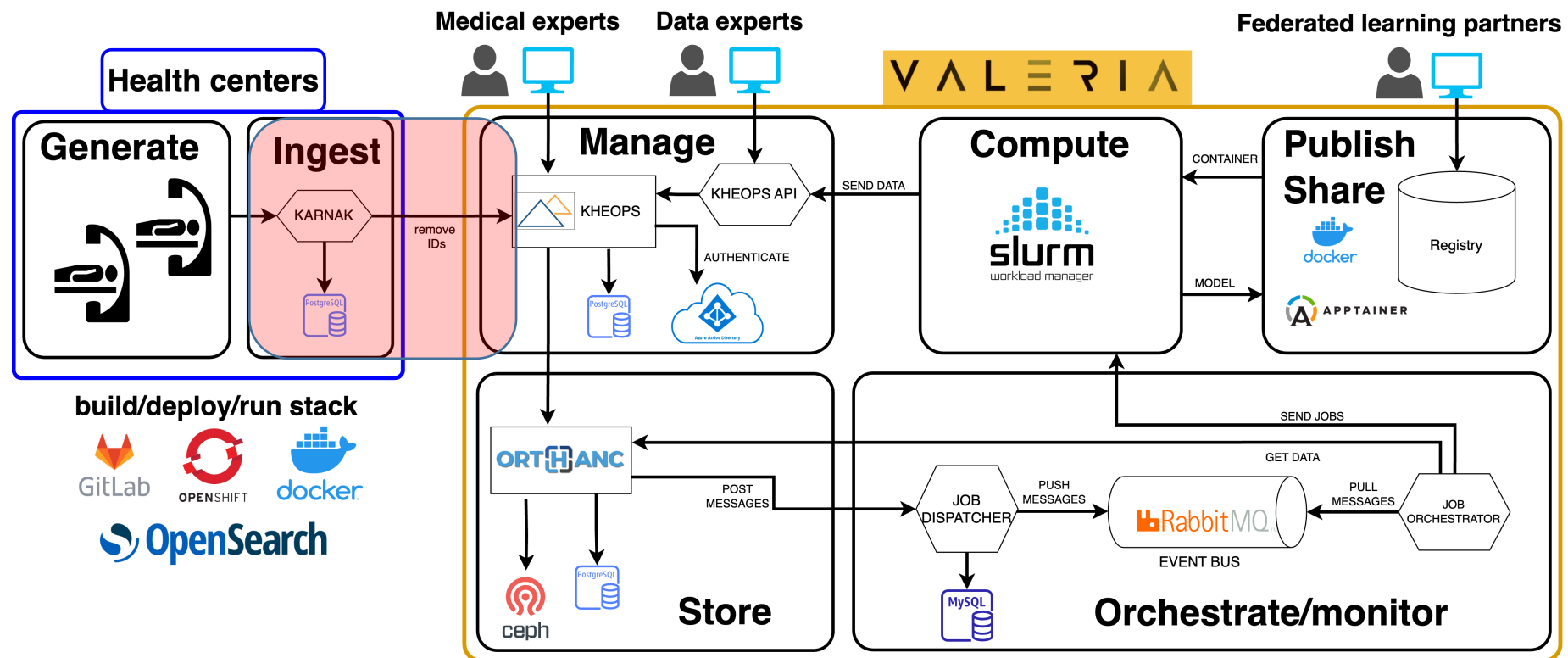
- Dedicated data management framework developed based on biobank philosophy: data are kept as long as there is scientific value
- Digital platform developed to manage IDs, de-identification, auth/autz, logging
- All data remain (mostly) on site



# Ethical / legal challenges

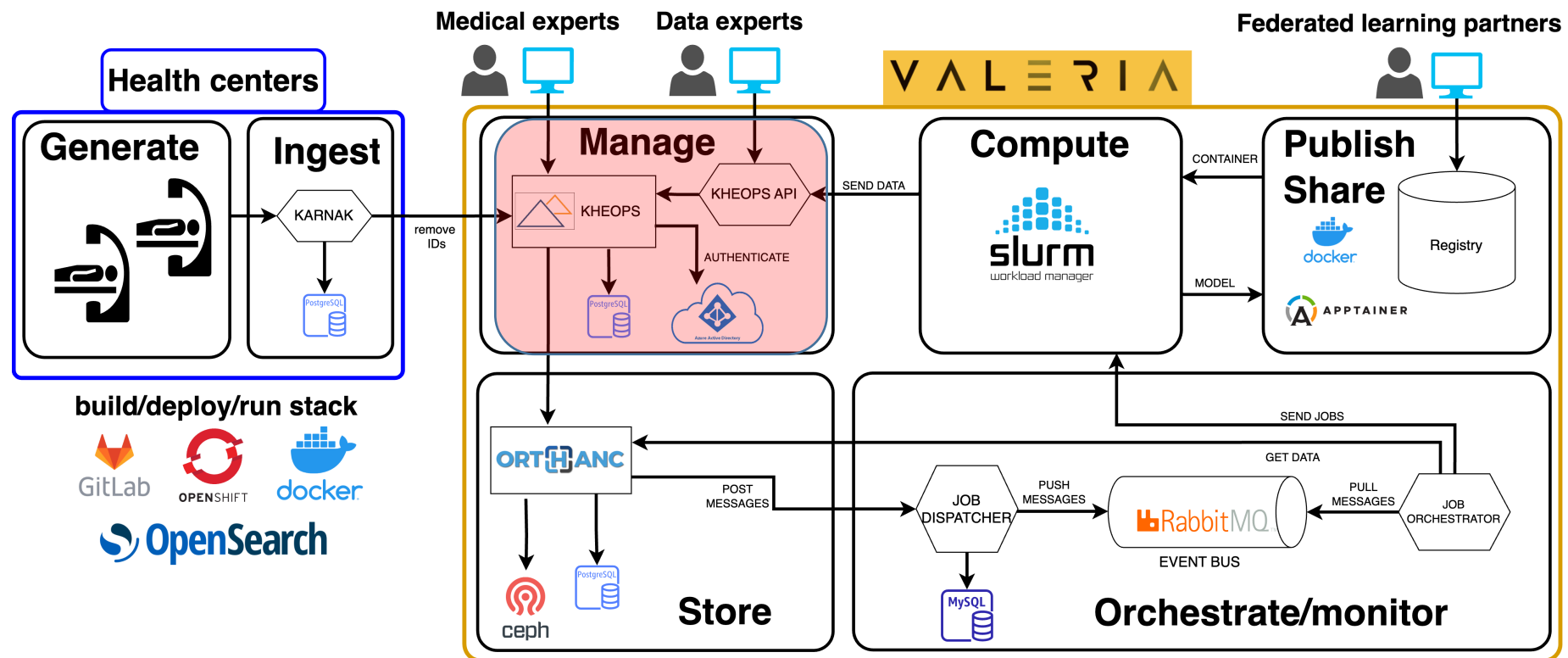


Images are de-identified according to specific profiles (DICOM or project-specific requirements) and ingested in PARADIM

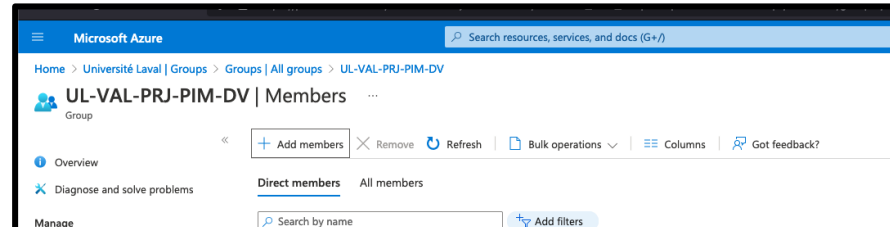


# Ethical / legal challenges

PARADIM translates our data management framework into a convenient solution (takes care of information governance)



# Ethical / legal challenges



## Benefits :

- Single sign-on (SSO): No user account and password management
- Requires an account at Université Laval: Knowledge of the user

Access reviews			
<input type="checkbox"/>		Samuel Ouellet	User
<input type="checkbox"/>		Sophie Tran Kiem	User
<input type="checkbox"/>		Yannick Lemarechal	User

Yannick Lemarechal Help ? Logout ↗ lang: en ▼

↻ 🔍

Last event Modalities

Sep 20th 2022 "CT", "SEG", "SR"



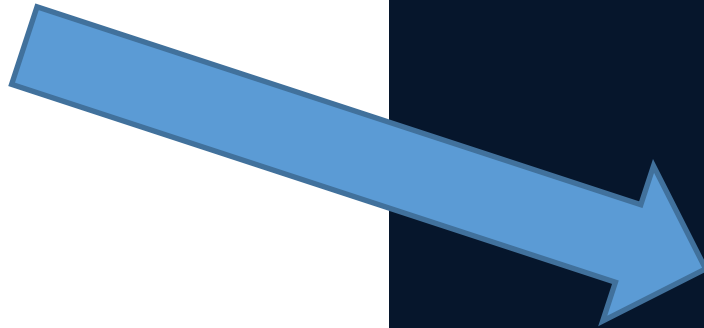
## UL-VAL-PRJ-PIM-DV

Désolé, nous rencontrons des problèmes pour vous connecter.

AADSTS50105: Your administrator has configured the application UL-VAL-PRJ-PIM-DV ('1a1d2774-eee1-4c31-ba8f-54cfeafa6e83') to block users unless they are specifically granted ('assigned') access to the application. The signed in user 'yannick.lemarechal@crchudequebec.ulaval.ca' is blocked because they are not a direct member of a group with access, nor had access directly assigned by an administrator. Please contact your administrator to assign access to this application.

# Ethical / legal challenges

Access and permissions are managed centrally



The screenshot displays the LIDC-IDRI web interface. At the top, there are navigation links for 'Inbox' and 'Albums'. The main header includes the 'LIDC-IDRI' logo and a star icon, along with 'Studies', 'Comments', and 'Settings' buttons. A sidebar on the left contains menu items: 'General', 'Users' (highlighted), 'Report', and 'Providers'. The main content area is titled 'List of users' and features an 'Invite a user' button. Below this, a list of users is shown with their email addresses and names in parentheses. At the bottom, the 'Album user settings' section is visible, listing several permissions with green checkmarks: 'Invite a user', 'Add studies / series', 'Remove studies / series', 'Show download button', 'Sharing', and 'Write comments'.

Album user settings	
<input checked="" type="checkbox"/> Invite a user	<input checked="" type="checkbox"/> Show download button
<input checked="" type="checkbox"/> Add studies / series	<input checked="" type="checkbox"/> Sharing
<input checked="" type="checkbox"/> Remove studies / series	<input checked="" type="checkbox"/> Write comments

# Ethical / legal challenges

User settings

General

Tokens

+ New token

Tokens  Show invalid tokens

Status	Description	Scope	Expiration date	Creation date	Last use date	Permission	
active	Access token for AI app	National Lung Screening Trial - NLST	Jun 4th 2023 16:20	May 4th 2023 16:21		read	Revoke
active	twitter_link	NSCLC	Jan 30th 2123 17:28	Jan 30th 2023 17:28		read, show download button	Revoke



Access for humans and machines through revokable tokens

# Data annotation

## Challenges

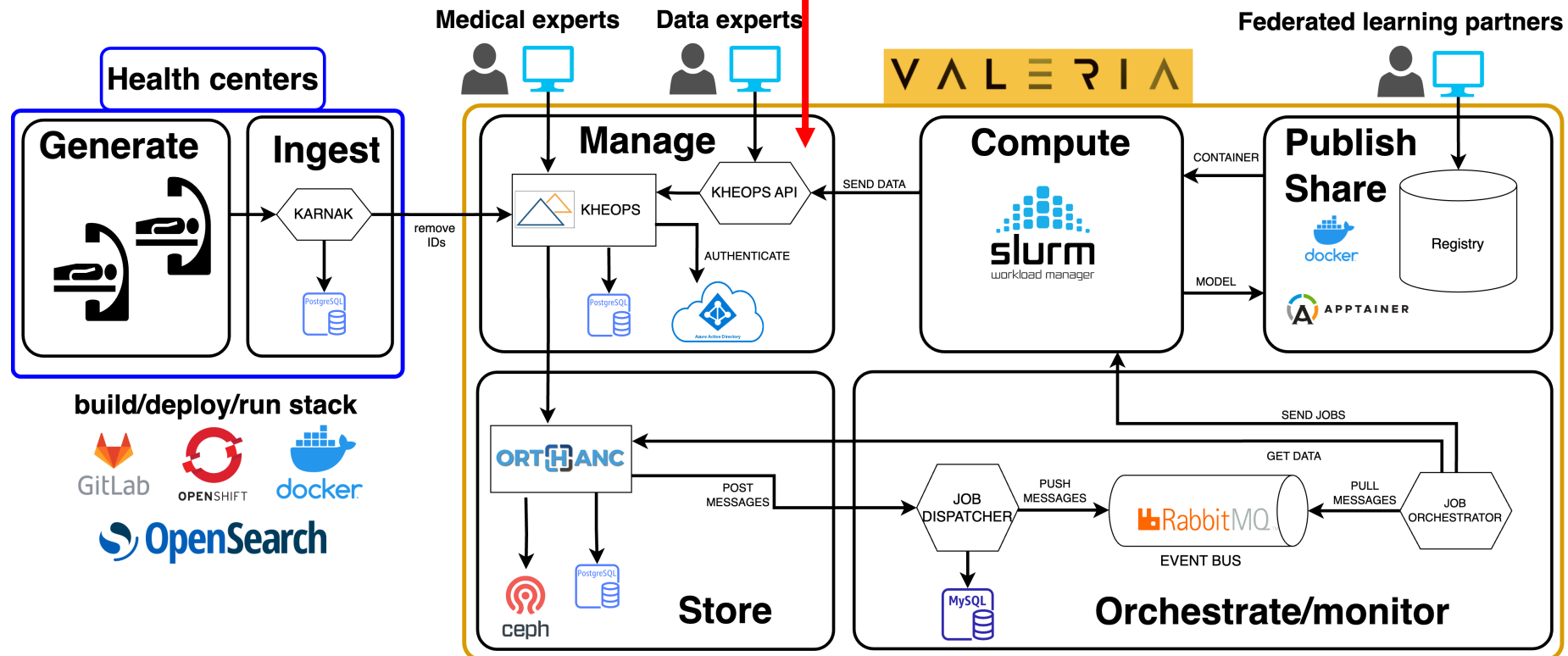
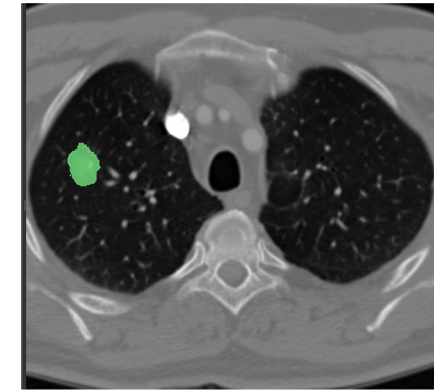
- Annotations required by supervised machine learning
- Information typically not preserved in a robust manner (XML, Excel file, proprietary formats)
- Context of annotation (who, when, how) typically not preserved
- Integration into clinical workflows is difficult

## Solution

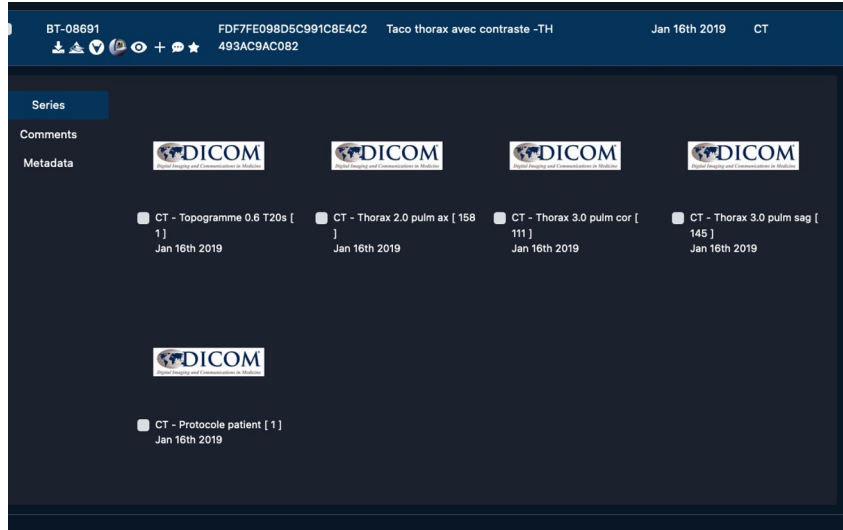
- An annotation pipeline was developed to address all challenges
- Annotations now preserved in an interoperable DICOM container which also keep context
- Multiuser support

# Data annotation

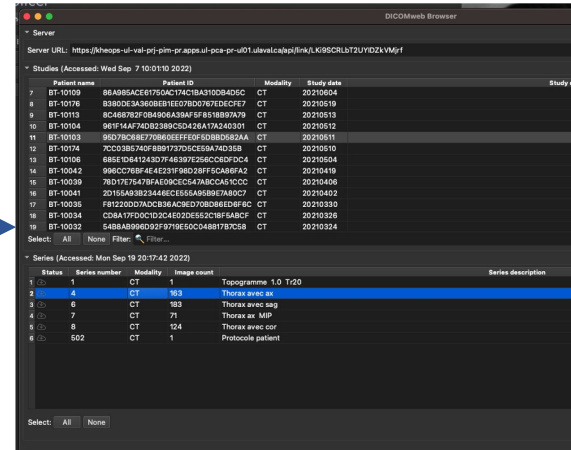
Third-party application can access albums through dicomweb protocol for manual annotation



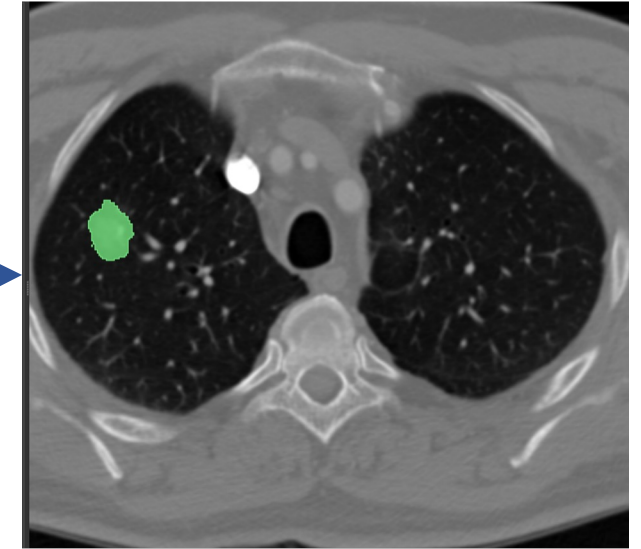
# Data annotation



CT Study

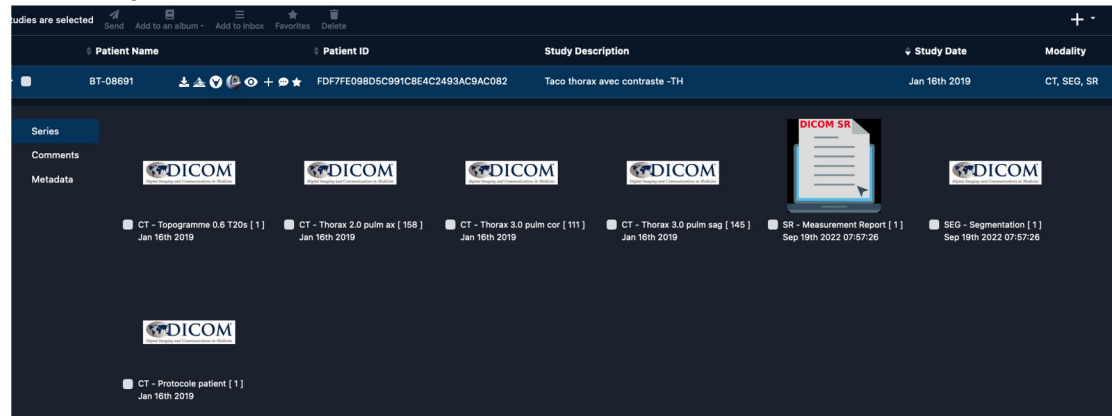


Data download

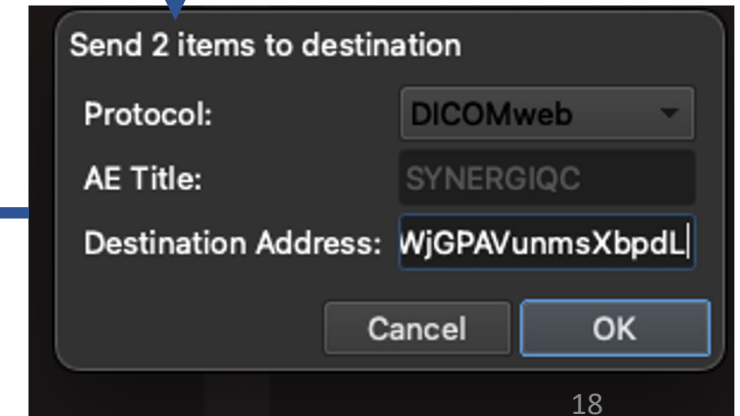


Annotation

## Study with CT exams and annotations

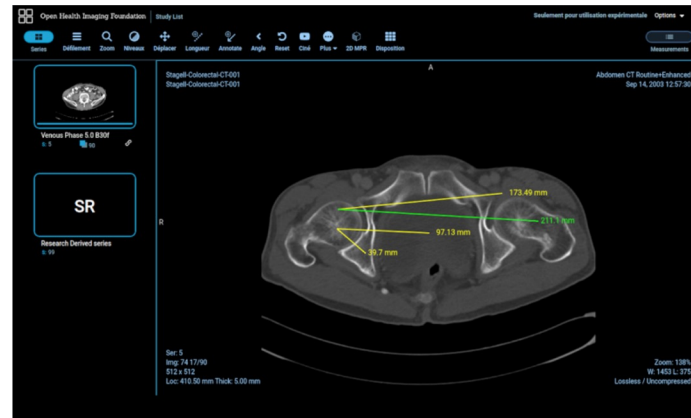


## Upload to PARADIM

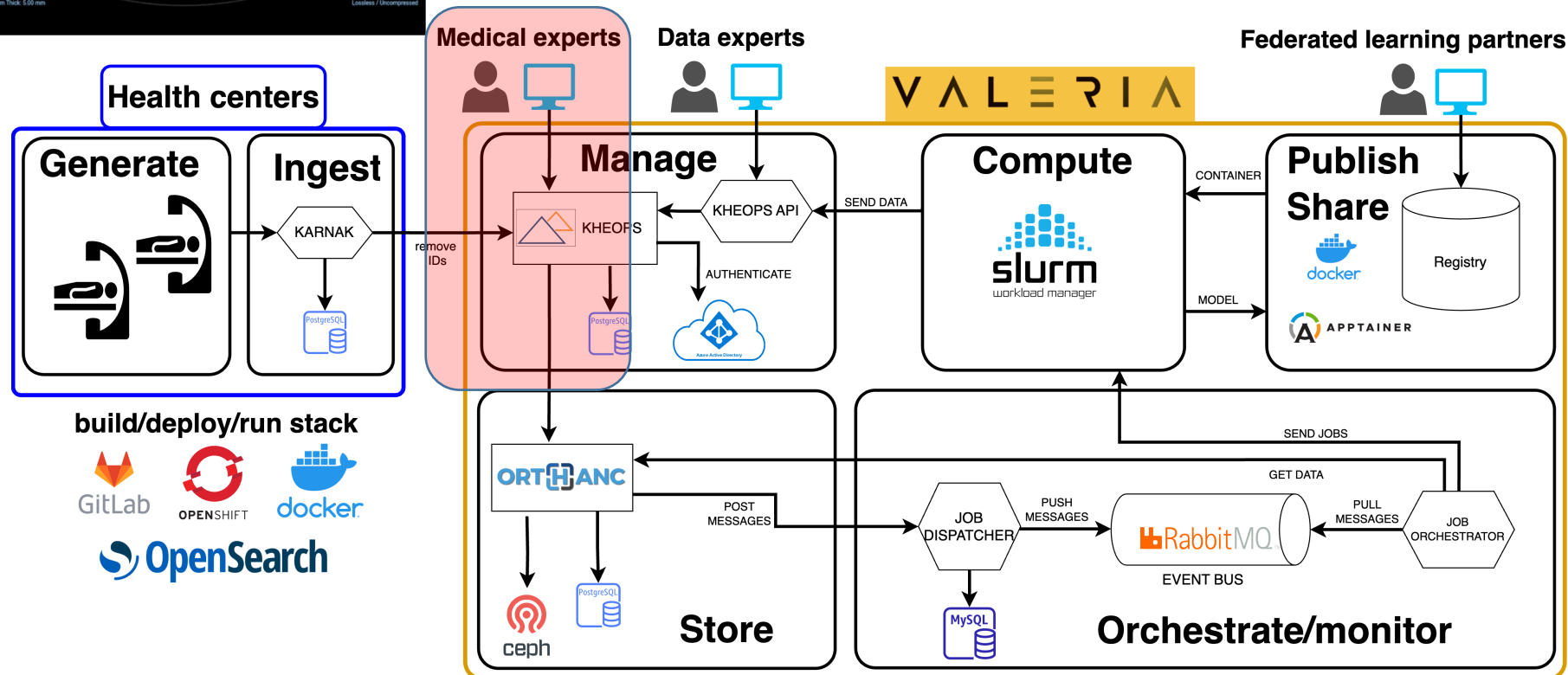




# Data annotation




Medical experts provide rich annotations (e.g. segmentations) that are kept in DICOM objects along with context (who, when, how)



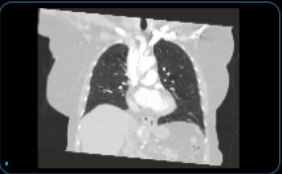
build/deploy/run stack  
 GitLab, OPENSIFT, docker, OpenSearch

# Data annotation

**Thorax sag**  
s: 6 176



**Thorax ax MIP**  
s: 7 171

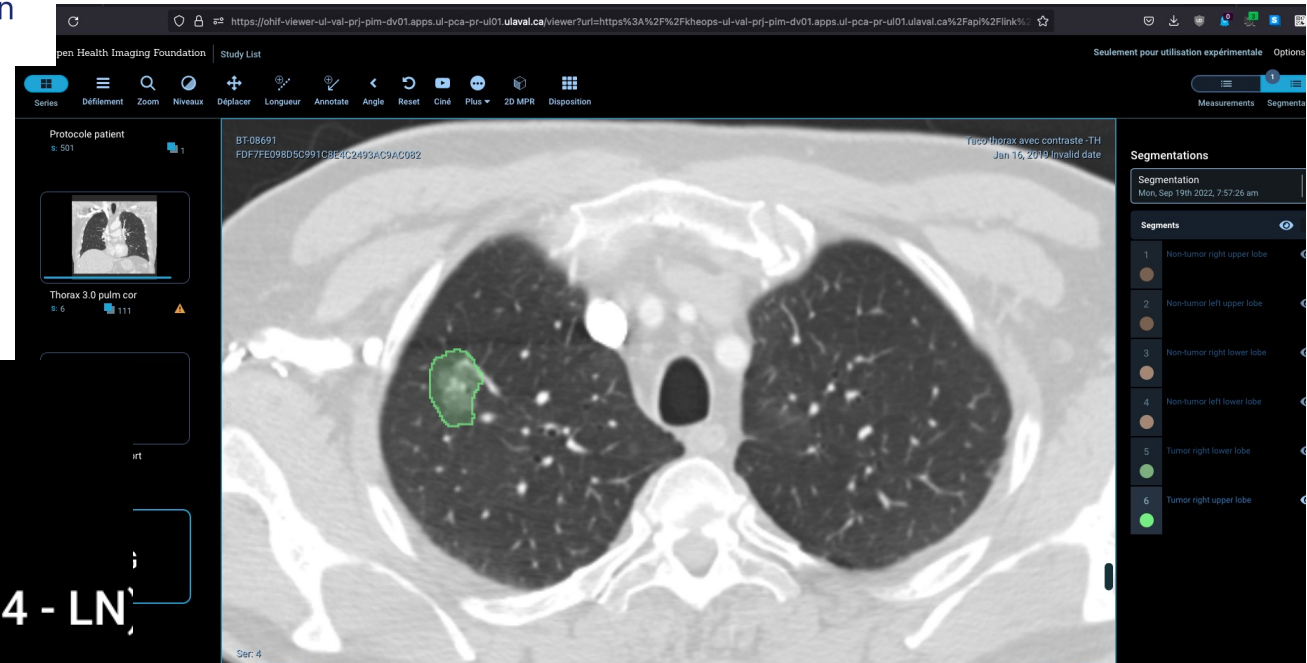


**Thorax cor**  
s: 8 123

**Patient:** BT-10271 (, #E2CA2A826989C094D25B97F44EE09D6E)  
**Study:** Taco thorax avec contraste -TH  
**Series:** undefined (#1023)  
**Manufacturer:** AGFA (undefined, #undefined)  
**Completion flag:** COMPLETE  
**Verification flag:** UNVERIFIED  
**Content Date/Time:** 20210131 081701

## Diagnostic Imaging Report (18748-4 - LN)

SpecificCharacterSet - ISO\_IR 100  
InstanceCreationDate - 20220901  
InstanceCreationTime - 171751.414986  
InstanceCreatorUID - 2.25.327946574954339906021728484647665005145  
SOPClassUID - 1.2.840.10008.5.1.4.1.1.88.11  
SOPInstanceUID - 2.25.19801702817787284799352855691316320403  
StudyDate - 20210302  
SeriesDate - 20210302  
ContentDate - 20210131  
StudyTime -  
ContentTime - 081701  
AccessionNumber -  
Modality - SR  
Manufacturer - AGFA  
ReferringPhysicianName -  
StudyDescription - Taco thorax avec contraste -TH  
ReferencedSOPInstanceUID - 2.25.14601284136870844387876283533888555749  
ReferencedSOPClassUID - 2.25.158755053078546251574755559975592455767  
ReferencedSOPInstanceUID - 2.25.19801702817787284799352855691316320403  
CodeValue - TH



# Volume

## Challenges

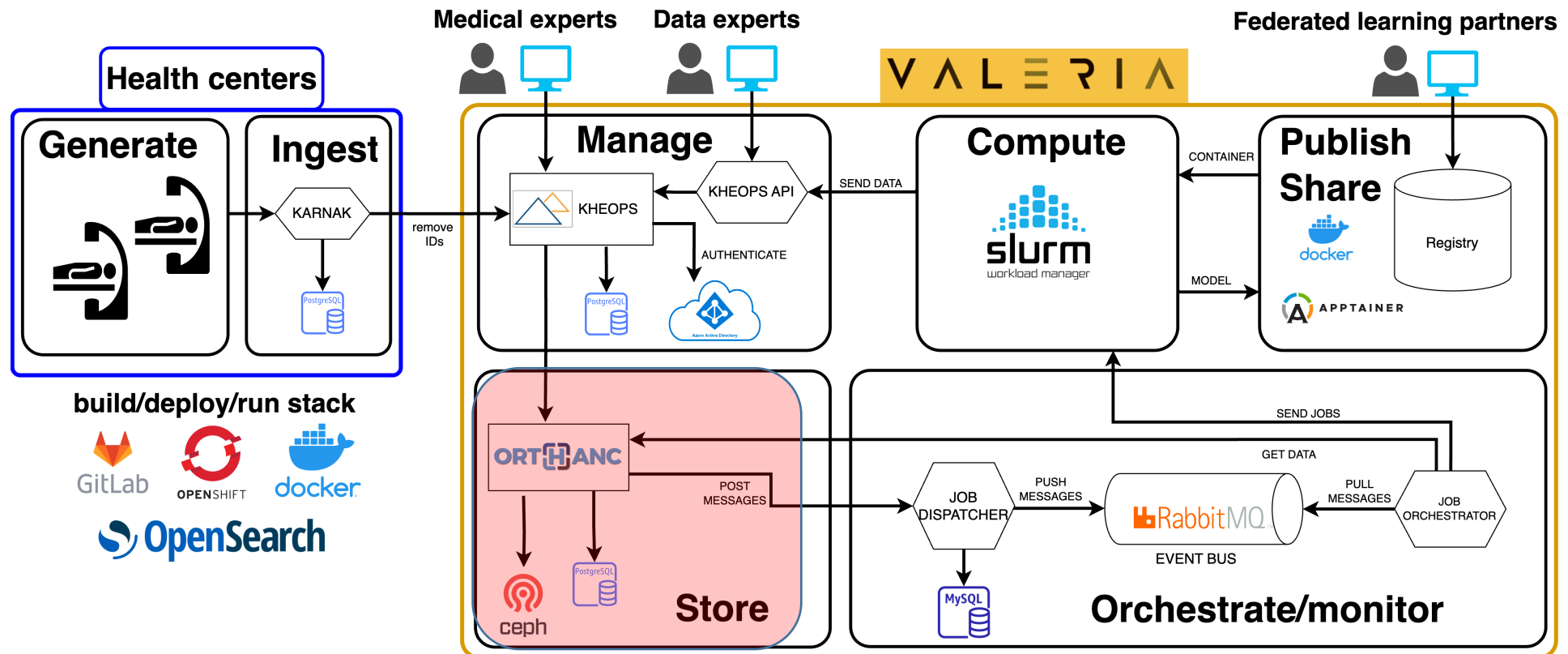
- Research in imaging requires significant resources (storage, compute)
- Finding previous studies is not trivial
- Querying local PACS can lead to disruptions in the clinic (filling the cache with old studies)
- Manual procedures/analyses not adapted to large volumes

## Solutions

- A DICOM-compliant open-source server was deployed (can initiate complex queries and balance the load)
- S3-compatible storage let envision future scaling up
- Triggered/scheduled-based job executor for automation

# Volume

DICOM-compliant backend + resilient storage (S3) ensures interoperability and robustness across the entire data lifecycle

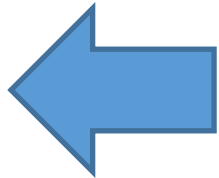


# Volume

resilient storage



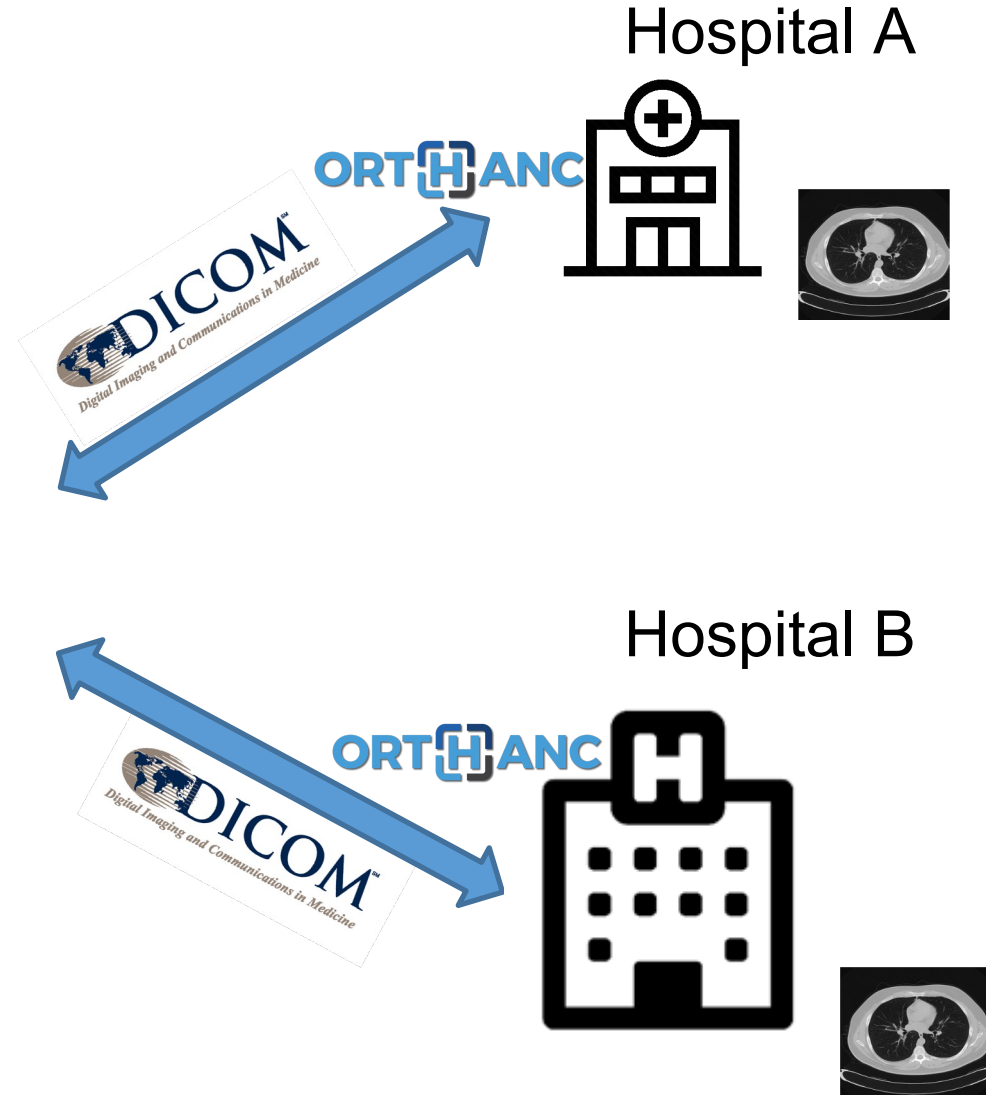
ceph



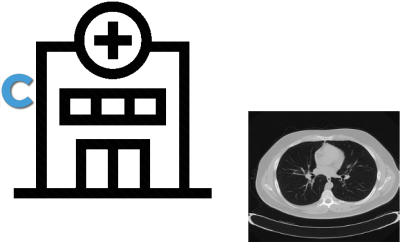
ORTHANC



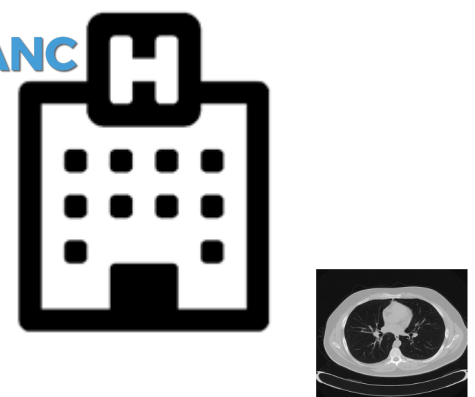
DICOM server acting as a swiss army knife to find/retrieve images



Hospital A



Hospital B



# Execution

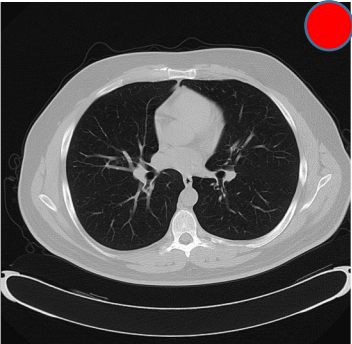
## Challenges

- Manual procedures/analyses not adapted to large volumes
- Traceability of operations typically not built-in

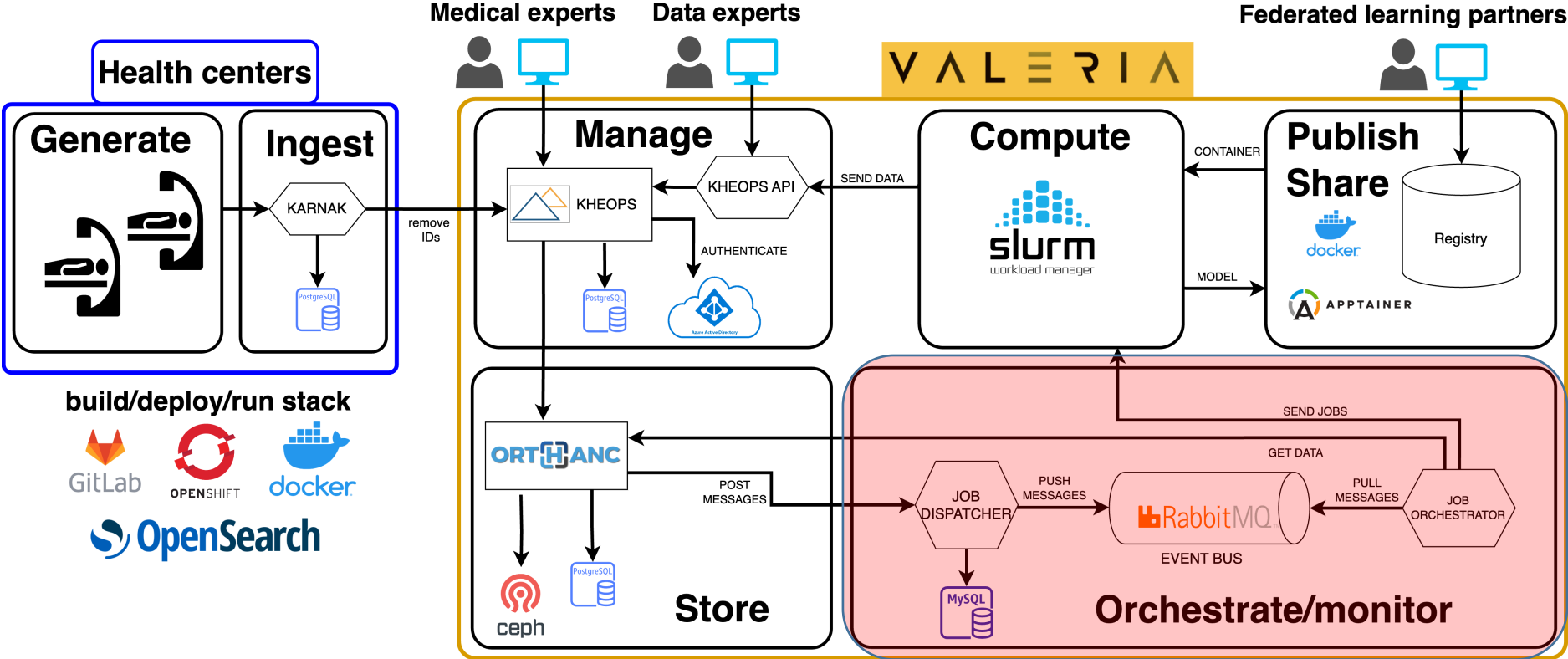
## Solution

- Triggered/scheduled based job executor
- A log report is associated with generated data
- Executables are embedded in versioned Docker images (no ambiguity on utilized version)

# Execution

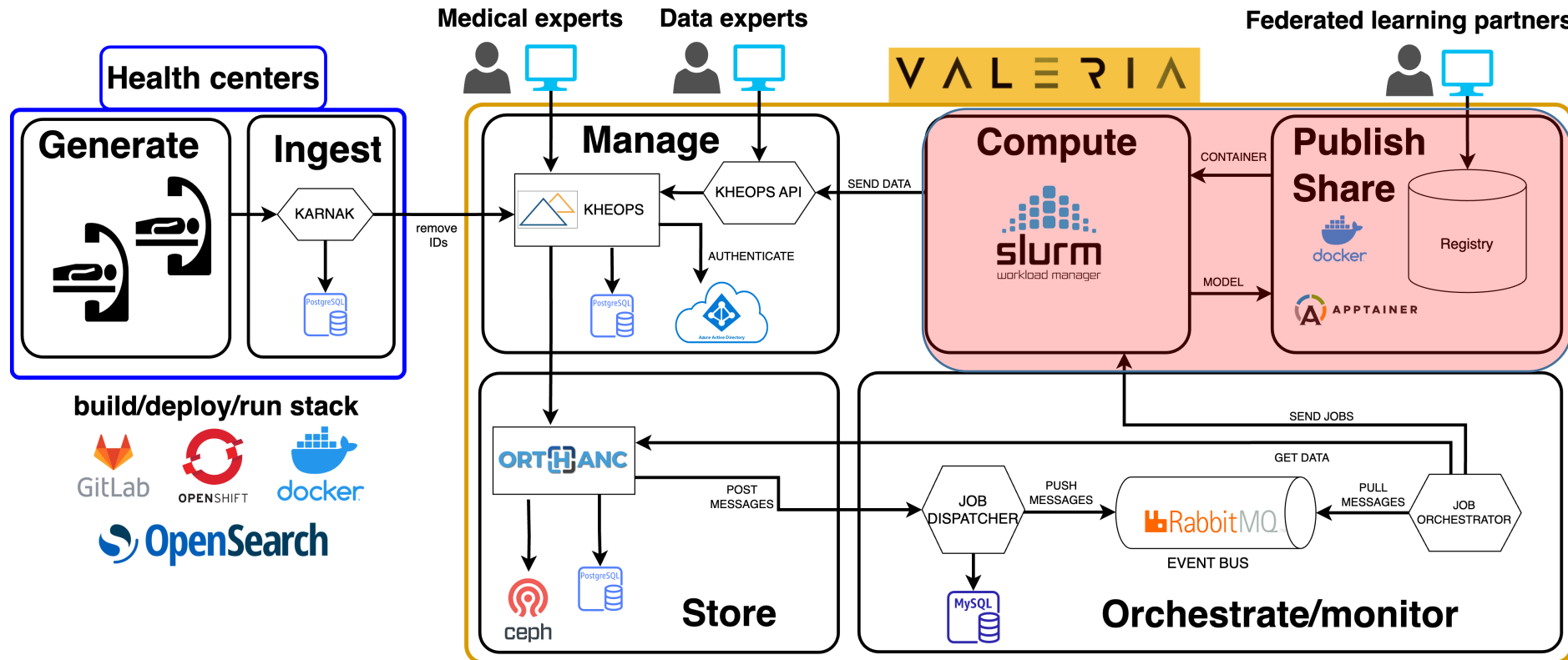
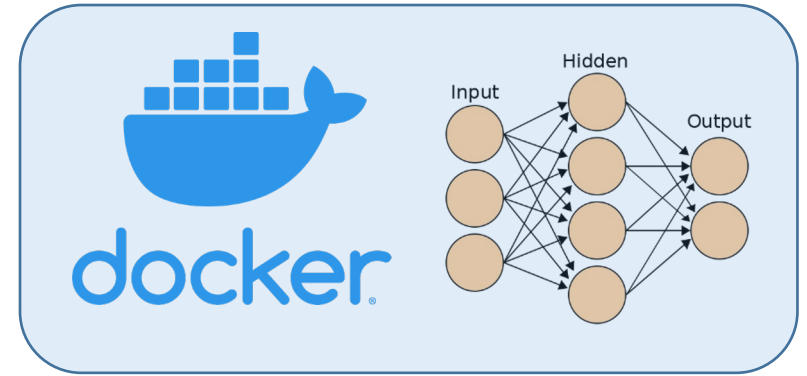


New images trigger calculation pipelines (retrain model, execute task) without manual intervention



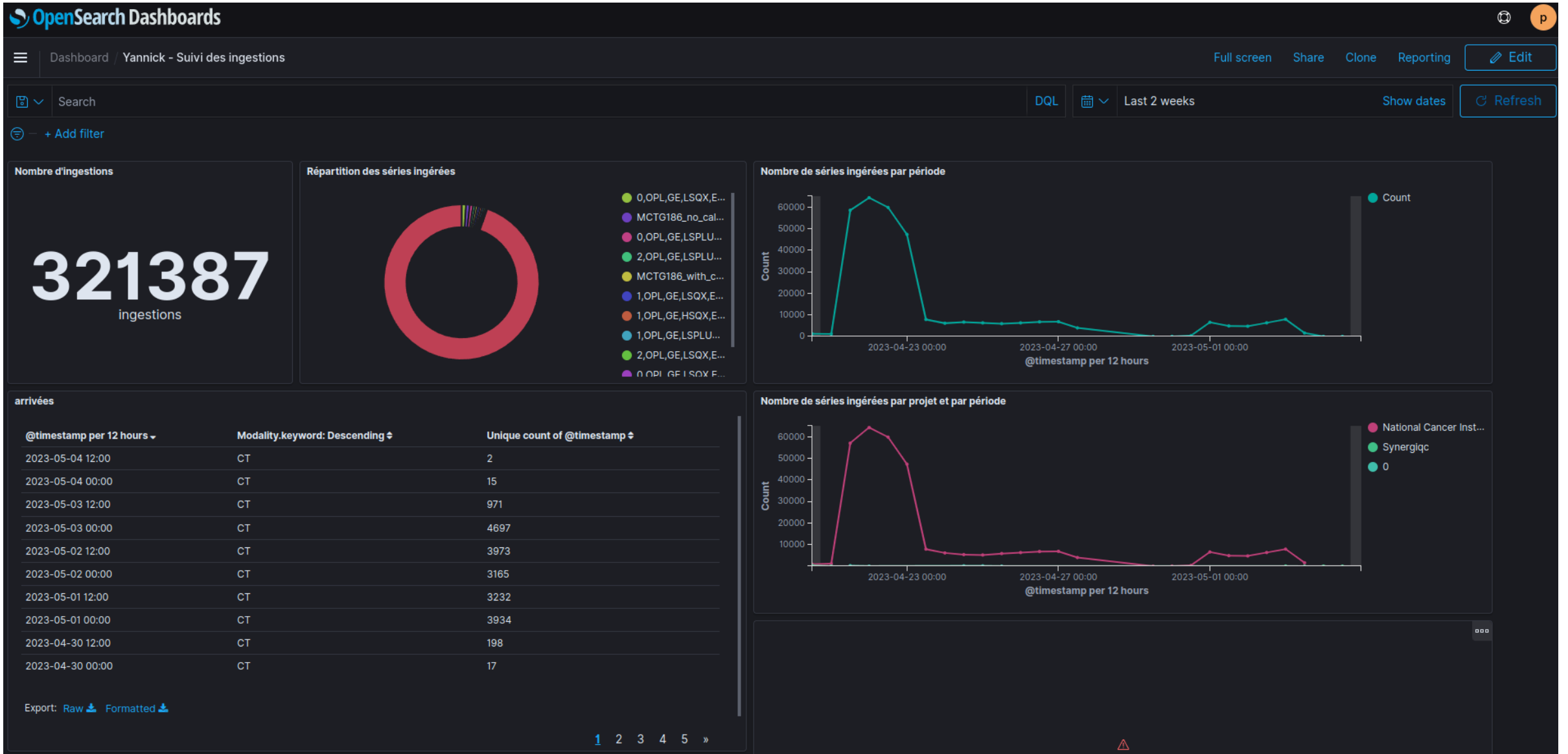
# Execution

AI models are managed as Docker images; data can be securely exposed to partners

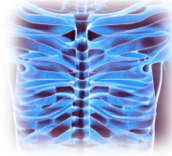




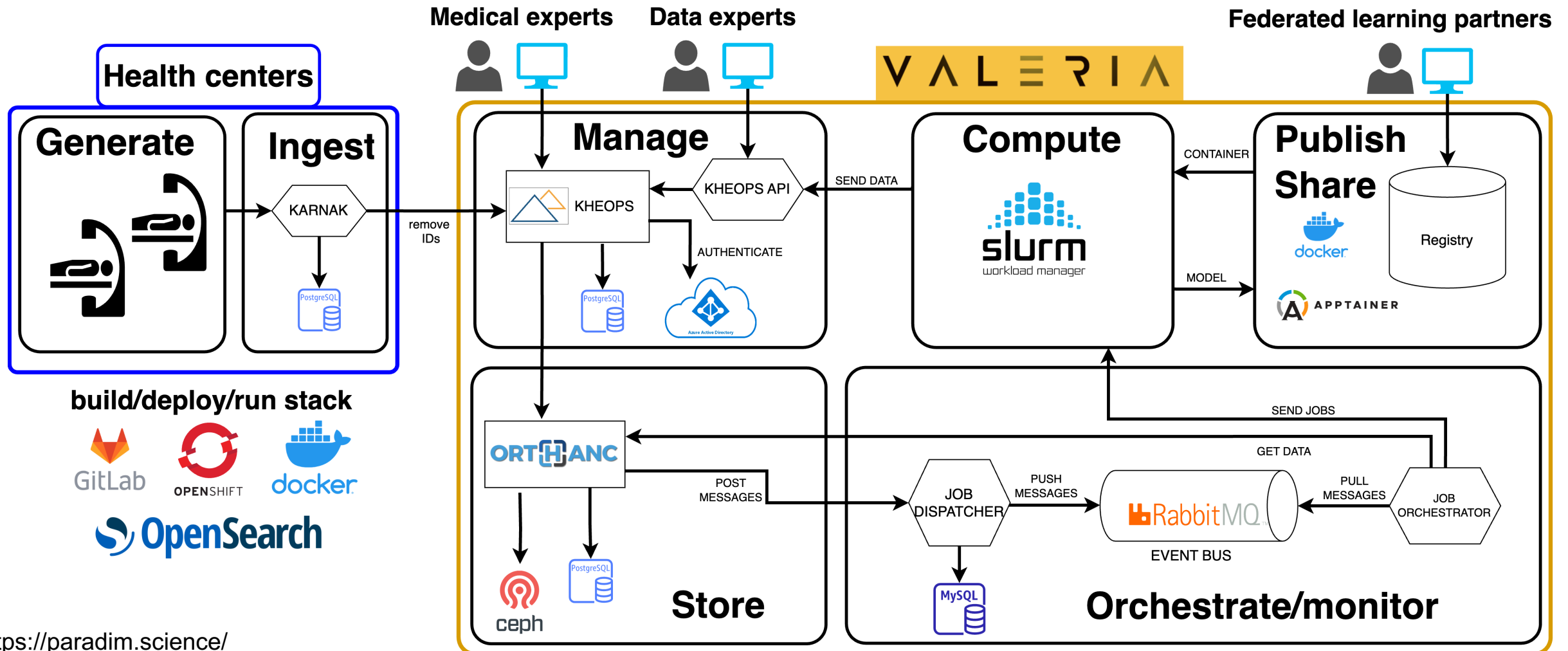
# Logging / monitoring



# The entire discovery cycle is industrialized



# PARADIM



# PARADIM: A digital infrastructure to support AI research in medical imaging

Yannick Lemaréchal,  
François Pelletier, Gabriel Couture, Pierre-Luc Asselin,  
Samuel Ouellet, Ronan Lefol, Luc Beaulieu  
and Philippe Després



CENTRE DE RECHERCHE  
INSTITUT UNIVERSITAIRE  
DE CARDIOLOGIE  
ET DE PNEUMOLOGIE  
DE QUÉBEC  
UNIVERSITÉ LAVAL

---