Using clinical images for machine learning research

the journey from hospital to HPC

Jonas Richiardi

@TranslationalML W3 unil.ch/tml

CHUV

UNIL | Université de Lausanne

TRANSLATIONAL MACHINE LEARNING LAB

Department of Radiology

[Sysang 1749]

A bird’s eye view of the journey

There is no escape!

Radiology

REVIEWS AND COMMENTARY • EDITORIAL

Machine Learning in Radiology: Resistance Is Futile

Mykol Larvie, MD, PhD

From the Division of Neuroradiology, Department of Nuclear Medicine, Cleveland Clinic, Imaging Institute, 9500 Euclid Ave, U15, Cleveland, OH 44195. Received October 4, 2018; revision requested October 9; final revision received October 9; accepted October 11. **Address correspondence to** the author (e-mail: *mlarvie@gmail.com*).

Conflicts of interest are listed at the end of this article.

See also the article by Ding et al in this issue.

Radiology 2019; 00:1–2 • https://doi.org/10.1148/radiol.2018182312 • Content code: NR •

> *"We are the Borg. Lower your shields and surrender your ships. We will add your biological and technological distinctiveness to our own. Your culture will adapt to service us. Resistance is futile." (1)*

There is a certain inevitability to the deployment of machine learning and other advanced computational methods in medicine. Especially in radiology, among the

A chief strength of the study is careful selection of the clinical problem. The best machine learning applications address "pain points"–problems that are not easily or well solved with existing resources, including a radiologist's skill. Accordingly, clinical interpretation of fluorodeoxyglucose (FDG) brain PET examinations is sorely in need of improvement. Reader performance is variable from cen-

ML is central for imaging biomarkers

Differential diagnosis

Intervention eligibility

Drug eligibility

Adverse effects

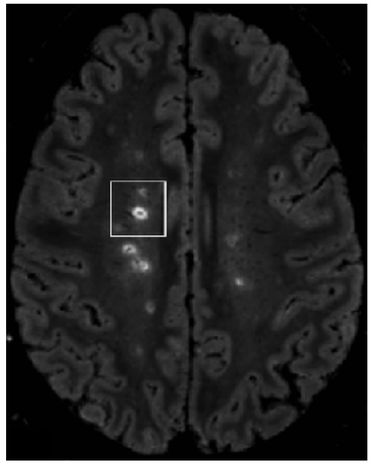
Drug response

…

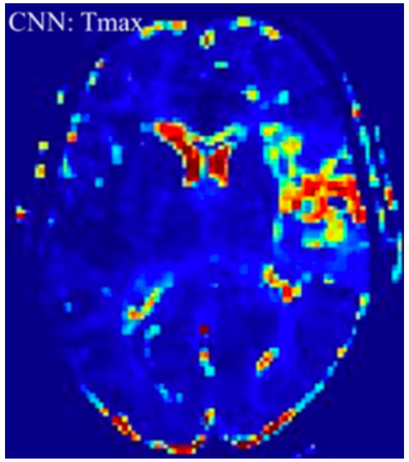
DOSAGE AND ADMINISTRATION

- Titration is required for treatment initiation. (2.1)
- The recommended maintenance dosage is 10 mg/kg administered as an intravenous infusion over approximately one hour every four weeks. (2.1)
- Obtain a recent (within one year) brain MRI prior to initiating treatment.

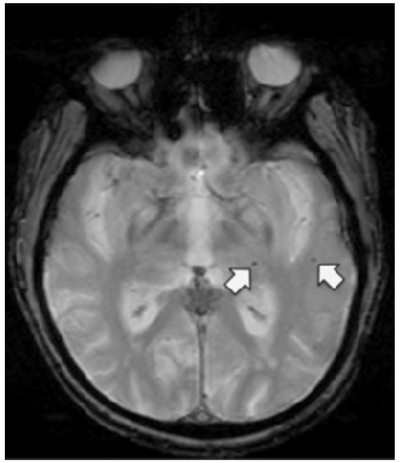
[Biogen 2021]



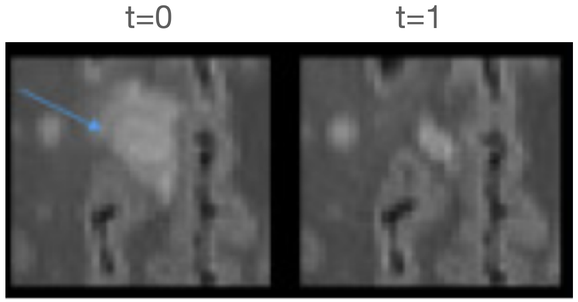
[Maggi, Fartaria, et al. 2020]



[Meier et al. 2020]



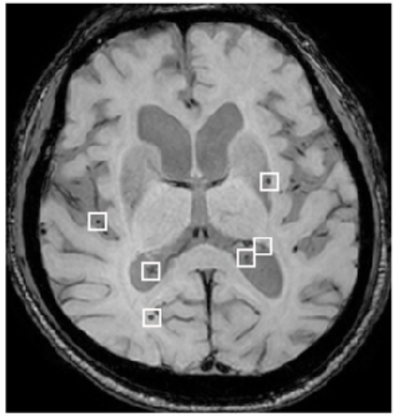
[Haller et al. 2013]



[Fartaria et al. 2019]

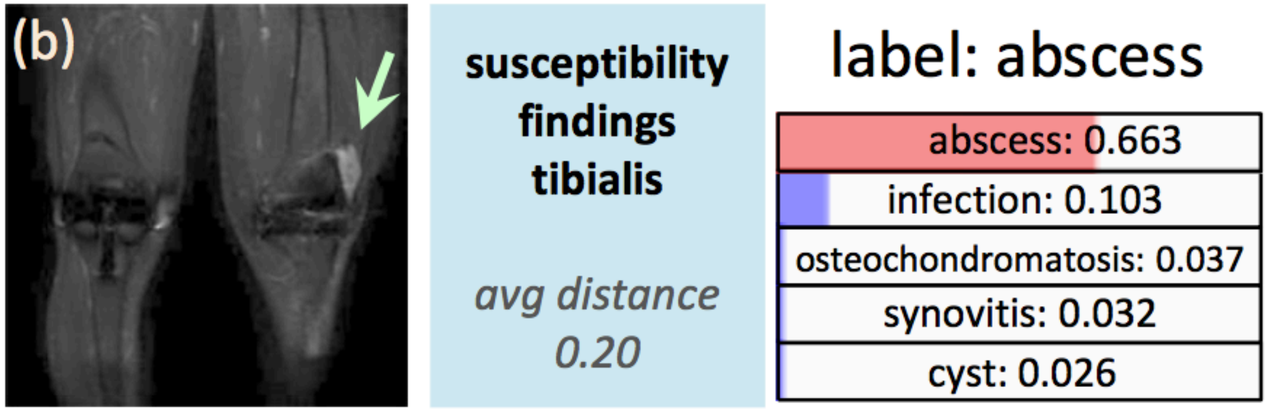
Two main ML tasks in radiology

Image → Image prediction



[Dou et al., IEEE TMI, 2016]

Image (+X) → Clinical prediction



[Shin et al., JMLR, 2016]

[Burn-Jones 1861]

Embarking on the journey

The map and the tools

Find and retrieve data

PACSMAN
Horus explorer (RDW)

Depersonalise data

Karnak

Store / share data

Karnak & Kheops

Annotate data

Mint
RepLab
RedCap

Develop & evaluate models

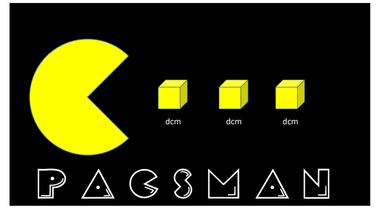
CHUV / UNIL HPC

Assess impact

radiology workflow
health economics
models

Finding images on the PACS at scale

Old way



Batch image query and retrieval tool

Known patient list

Here's a list of 500 patients and exam dates, give me all their MR and SR images

I need to type them in one by one, come back in a week or two and bring a hard disk

PatientID	StudyDate	Modality
12345	20200117	MR
12345	20200117	SR
8888	20180915	MR
…	…	…

Unknown patient list

How many images with ProtocolName 'WIP748' were acquired between 2018 and 2020?

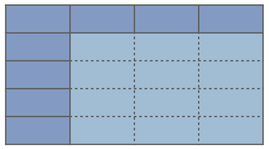
Sorry, nobody will ever know

ProtocolName	StudyDate	Modality
WIP748	20180101-20201231	MR

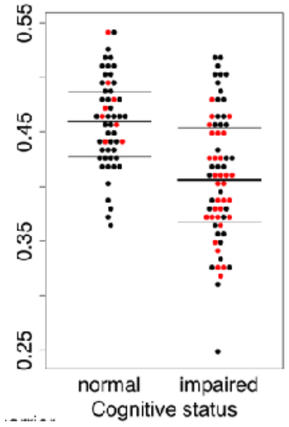
With grateful acknowledgement of contributions from Frédéric Pedron (RAD), Firas Ben Othman (RAD), Seb Tourbier (RAD), Roger Schaer (HES-SO), Alex Wetzel (DSI)

Finding other medical data

clinical scores

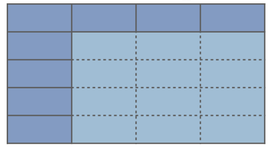


Neuropsy

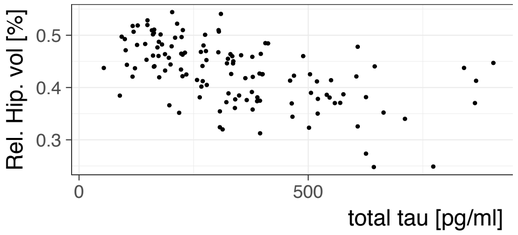


5-100

assays

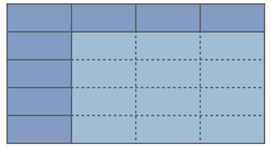


CSF Amyloid

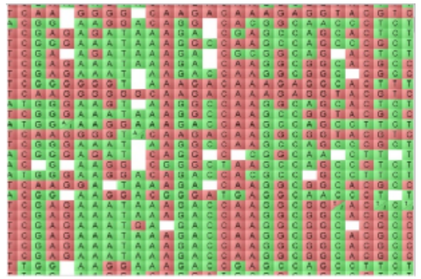


1-5

genetics



SNPs



1-10M

EHRs

Rad. report

100-600

Horus explorer / Research data warehouse / custom RedCap (e.g. RegCOVID)

Depersonalising DICOM images

Tag (field) level

Image Metadata:

Key	Value
Specific Character Set	ISO 2022 IR 100
Image Type	ORIGINAL\PRIMARY\M\NORM\DIS2D\MFSPLIT
Instance Creation Date	20210728
Instance Creation Time	[redacted]
SOP Class UID	1.2.840.10008.5.1.4.1.1.4
SOP Instance UID	1.2.840.1[redacted]
Study Date	20210728
Series Date	20210728
Acquisition Date	20210728
Content Date	20210728
Acquisition Date Time	[redacted]
Study Time	[redacted]
Series Time	[redacted]
Acquisition Time	[redacted]
Content Time	[redacted]
Accession Number	
Modality	MR
Modalities in Study	MR\SR\PR
Manufacturer	Siemens

basic approach - rule-based

	blank	replace	modify
names		X	
unique IDs		X	
dates			X

but….

`PatientMotherBirthName (0x0010,0x1060)`

undocumented private tags

pixel data can still hold personal information

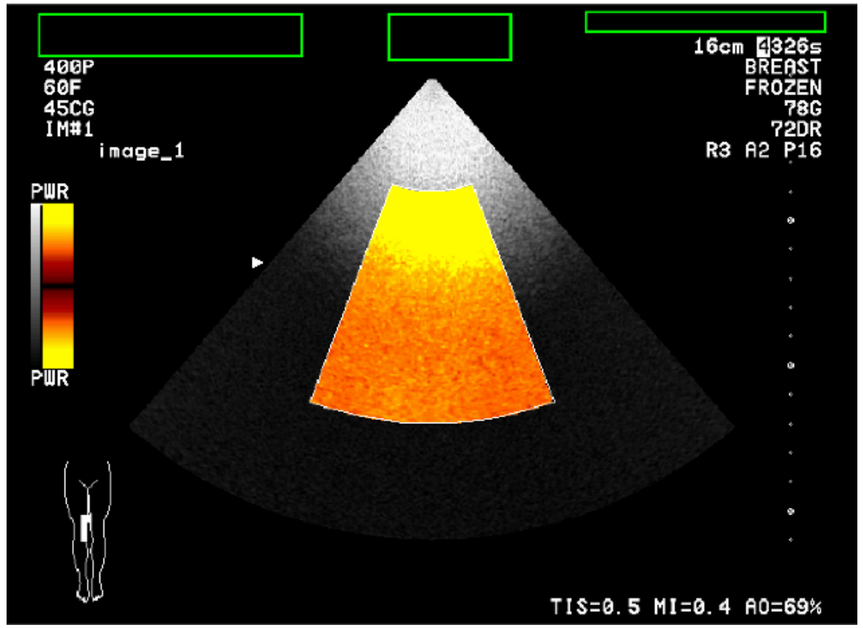
Depersonalising DICOM images

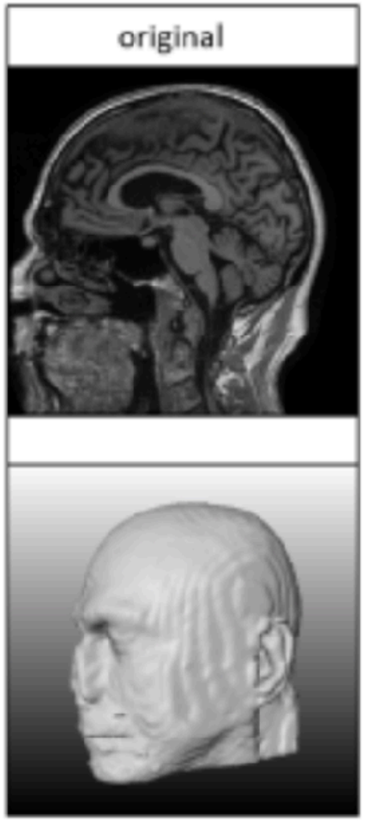
Image level

Delete

Mask

Deface/reface





[dclunie.com]

[Huelnhagen et al., ISMRM 2021]

Depersonalising DICOM data

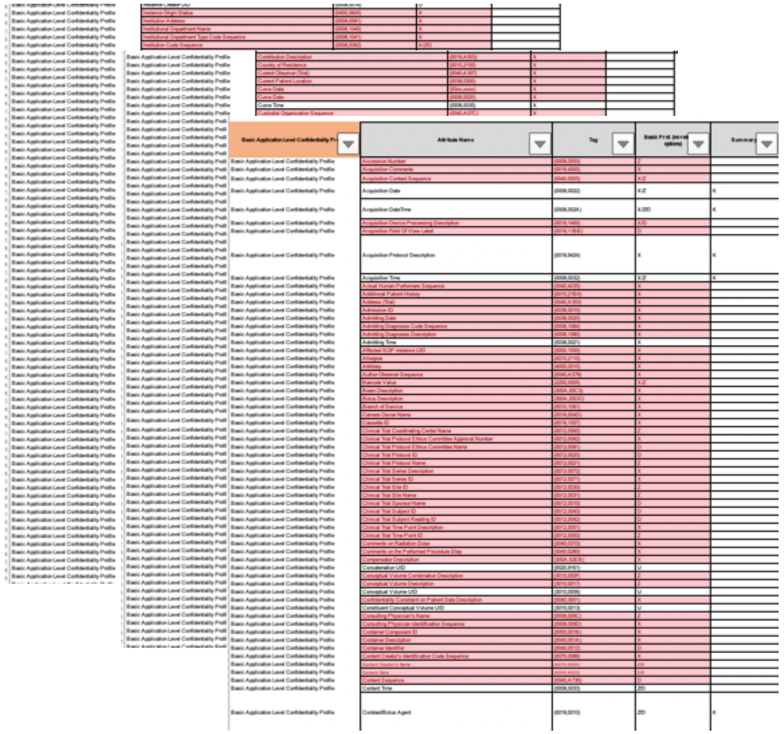
Towards standardization: SPHN DCC working Group de-ID efforts

Risk-based approach

Risk Profile					
General Questions			**Risk value subtotal**	**Category weight**	**Risk score**
No. of high risk answers	4		254	35%	2
BioMedIT usage	Yes				
Structured Data					
No. of high risk rules selected:	2		0	15%	0
Unstructured Data					
No. of high risk rules selected:	0		0	20%	0
Multimedia Data					
No. of high risk rules selected:	0		0	10%	0
Omics Data					
No. of high risk answers:	0		0	20%	0
Risk assessment outcome					
No. of high risk options selected:	6		Total Risk Score:		0.23

[SPHN DCC de-ID WG]

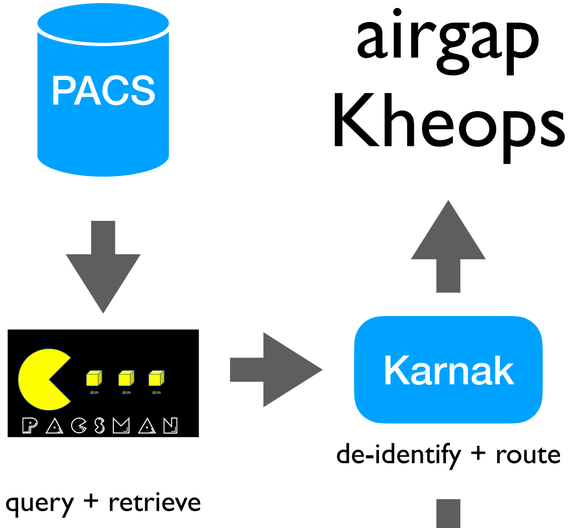
Exhaustive tag ruleset



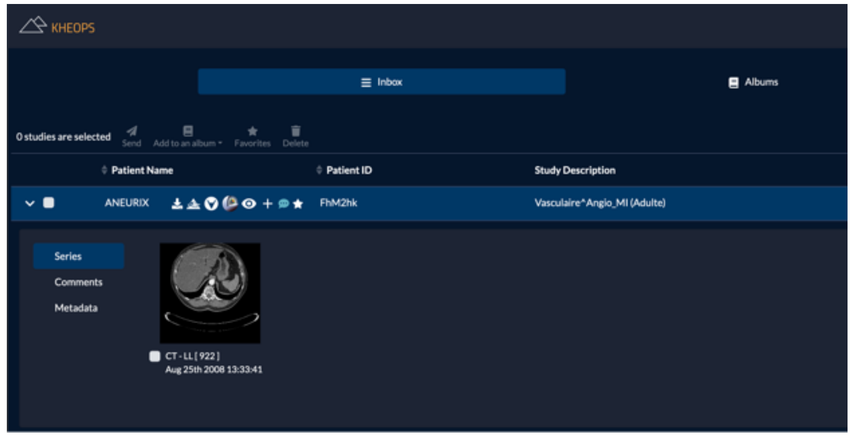
[SPHN DCC de-ID WG]

Storing and sharing DICOM data

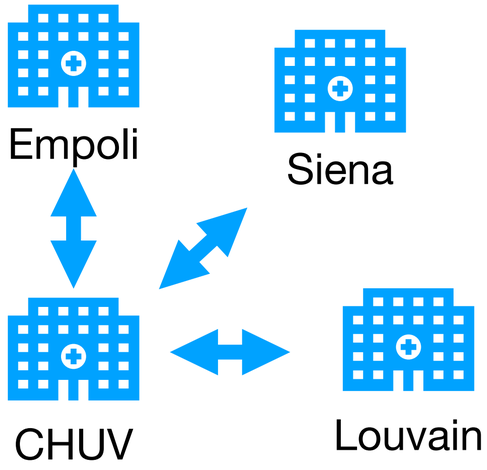
kheops-rg.chuv.ch



kheopspub.unil.ch



example: multicentric COVID CXR (Catherine Beigelman)



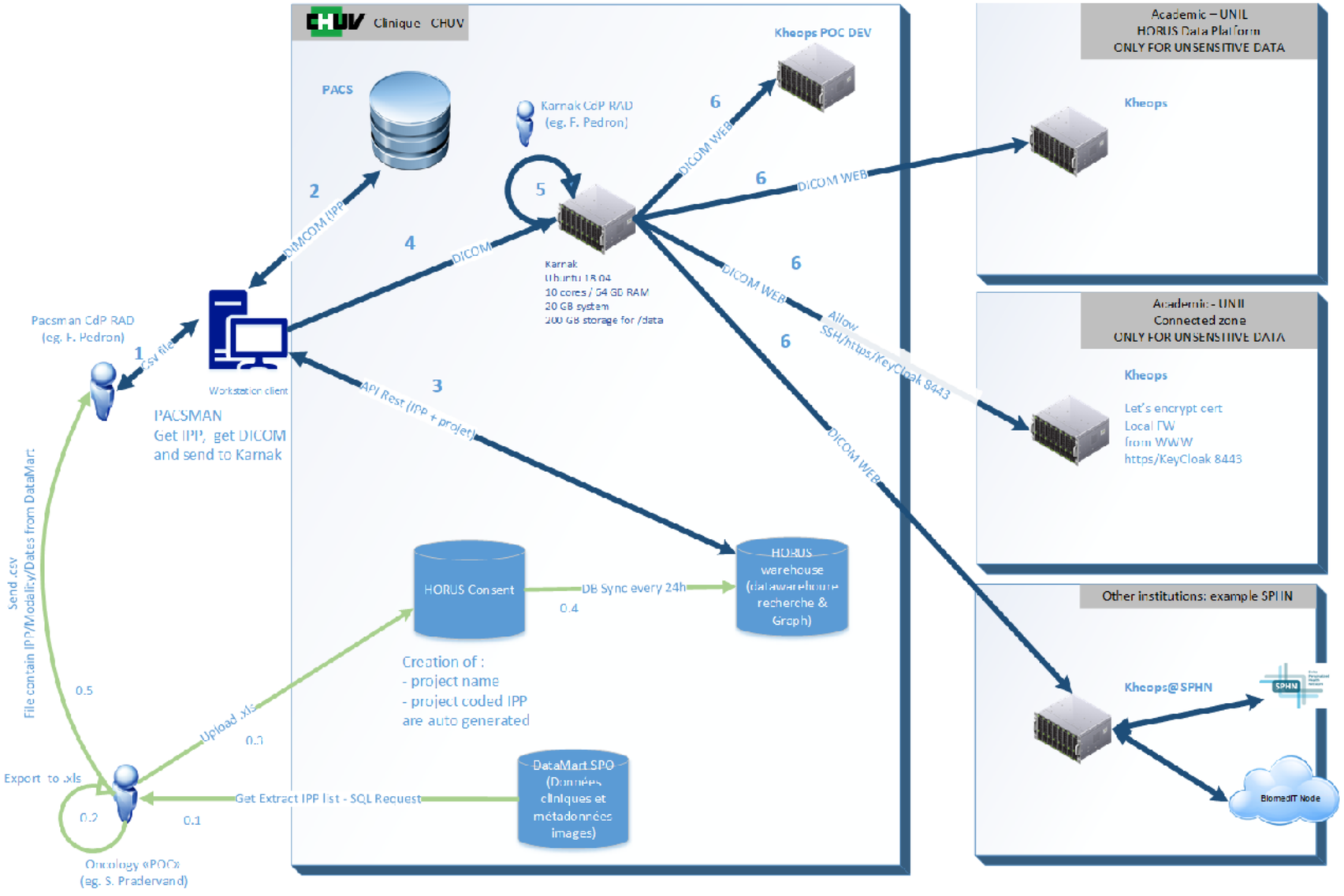
Collaboration with Frédéric Pedron (CHUV/RAD), Alexandre Wetzel (CHUV/DSI), Yves Jaggi (CHUV(DSI), Solange Zoergibel (CHUV/DSI), Roberto Fabbretti (UNIL/DCSR), Roger Schaer (HES-SO), Adrien Depeursinge (CHUV/NUC+HES-SO)

Complete DICOM depersonalisation workflow

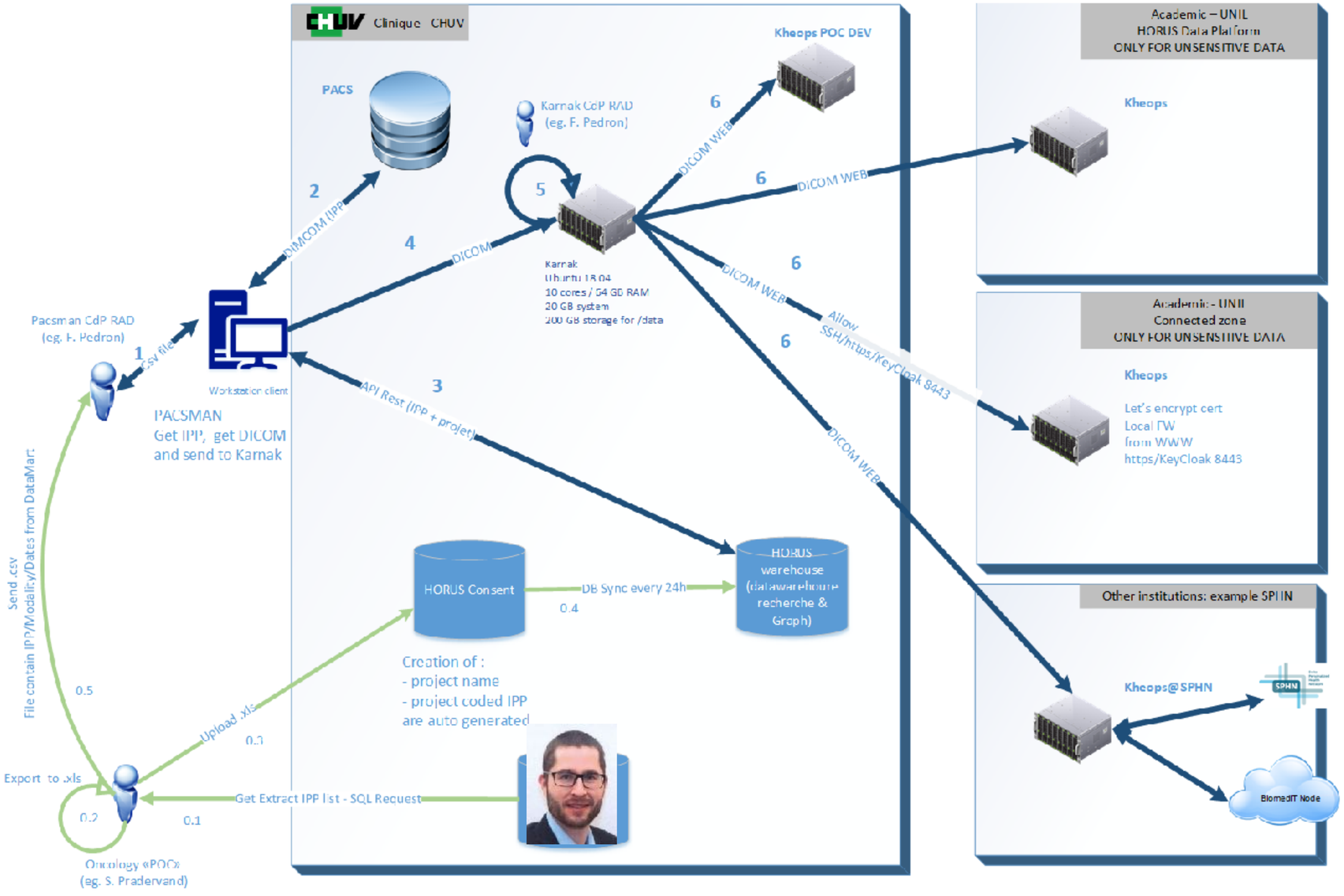
The CHUV approach

Depersonalisation is great,
but we need consistent IDs
and timings across images
and clinical data

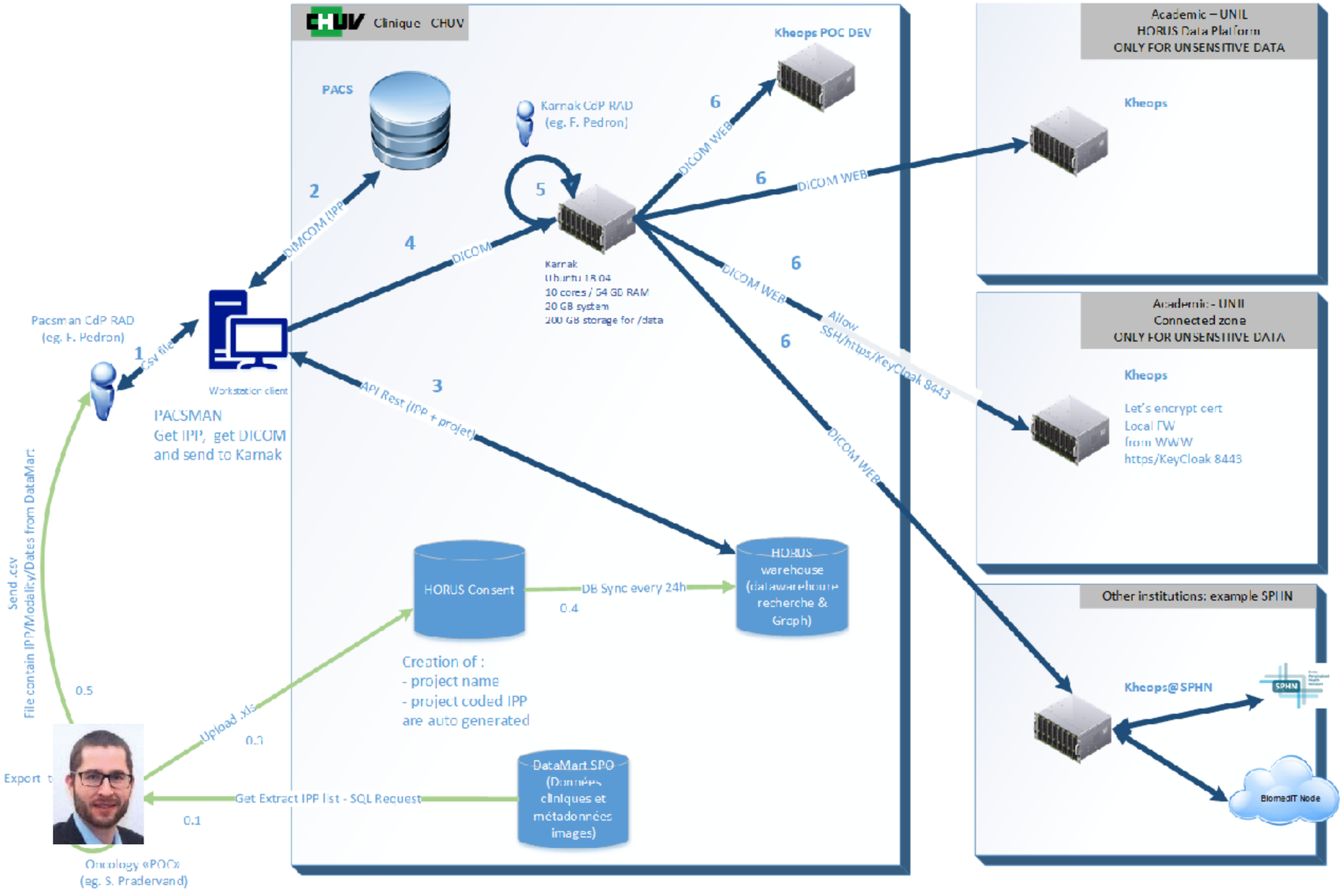
HORUS Image


Clinique CHUV
PACS
Karnak CdP RAD
(eg. F. Pedron)
Kheops POC DEV
Karnak
Ubuntu 18.04
10 cores / 64 GB RAM
20 GB system
200 GB storage for /data
DICOM (IPP)
DICOM
DICOM WEB
API Rest (IPP + project)
Allow SSH/https/KeyCloak 8443
Pacsman CdP RAD
(eg. F. Pedron)
Workstation client
PACSMAN
Get IPP, get DICOM
and send to Karnak
Send .csv
File contain IPP/Modality/Dates from DataMart
HORUS Consent
DB Sync every 24h
HORUS warehouse (datawarehouse recherche & Graph)
Creation of :
- project name
- project coded IPP
are auto generated
Upload .xls
Export to .xls
Get Extract IPP list - SQL Request
DataMart SPO (Données cliniques et métadonnées images)
Oncology «POC»
(eg. S. Pradervand)
Academic – UNIL
HORUS Data Platform
ONLY FOR UNSENSITIVE DATA
Kheops
Academic - UNIL
Connected zone
ONLY FOR UNSENSITIVE DATA
Kheops
Let's encrypt cert
Local FW
from WWW
https/KeyCloak 8443
Other institutions: example SPHN
Kheops@SPHN
SPHN
BiomedIT Node

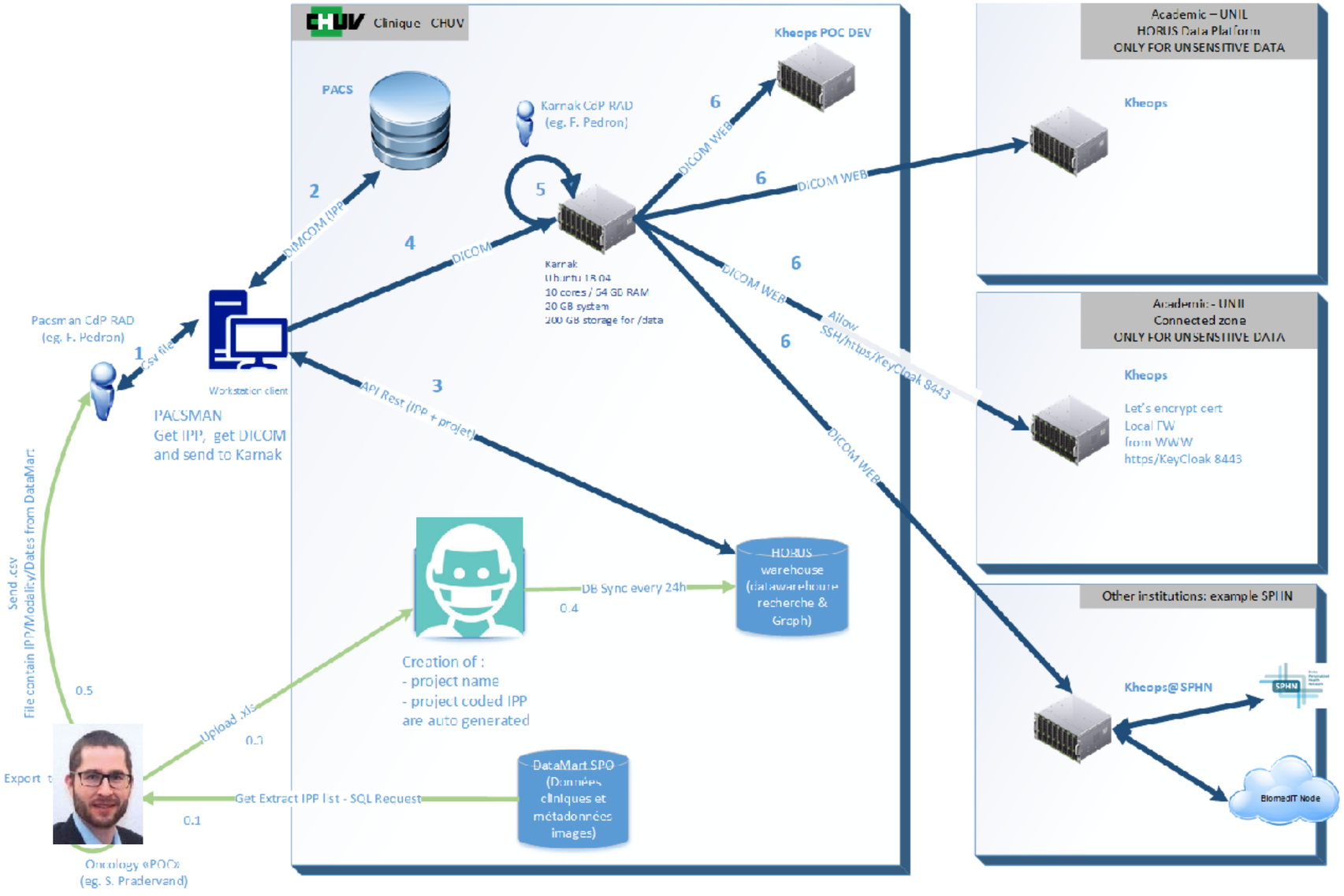
HORUS Image


Clinique CHUV
PACS
Karnak CdP RAD
(eg. F. Pedron)
Kheops POC DEV
Karnak
Ubuntu 18.04
10 cores / 64 GB RAM
20 GB system
200 GB storage for /data
Pacsman CdP RAD
(eg. F. Pedron)
Workstation client
PACSMAN
Get IPP, get DICOM
and send to Karnak
DIMCOM (IPP)
DICOM
API Rest (IPP + project)
DICOM WEB
Allow SSH/https/KeyCloak 8443
Academic – UNIL
HORUS Data Platform
ONLY FOR UNSENSITIVE DATA
Kheops
Academic - UNIL
Connected zone
ONLY FOR UNSENSITIVE DATA
Kheops
Let's encrypt cert
Local FW
from WWW
https/KeyCloak 8443
Other institutions: example SPHN
Kheops@SPHN
SPHN
BiomedIT Node
HORUS Consent
DB Sync every 24h
HORUS
warehouse
(datawarehouse
recherche &
Graph)
Creation of :
- project name
- project coded IPP
are auto generated
Send .csv
File contain IPP/Modality/Dates from DataMart
Upload .xls
Export to .xls
Get Extract IPP list - SQL Request
Oncology «POC»
(eg. S. Pradervand)

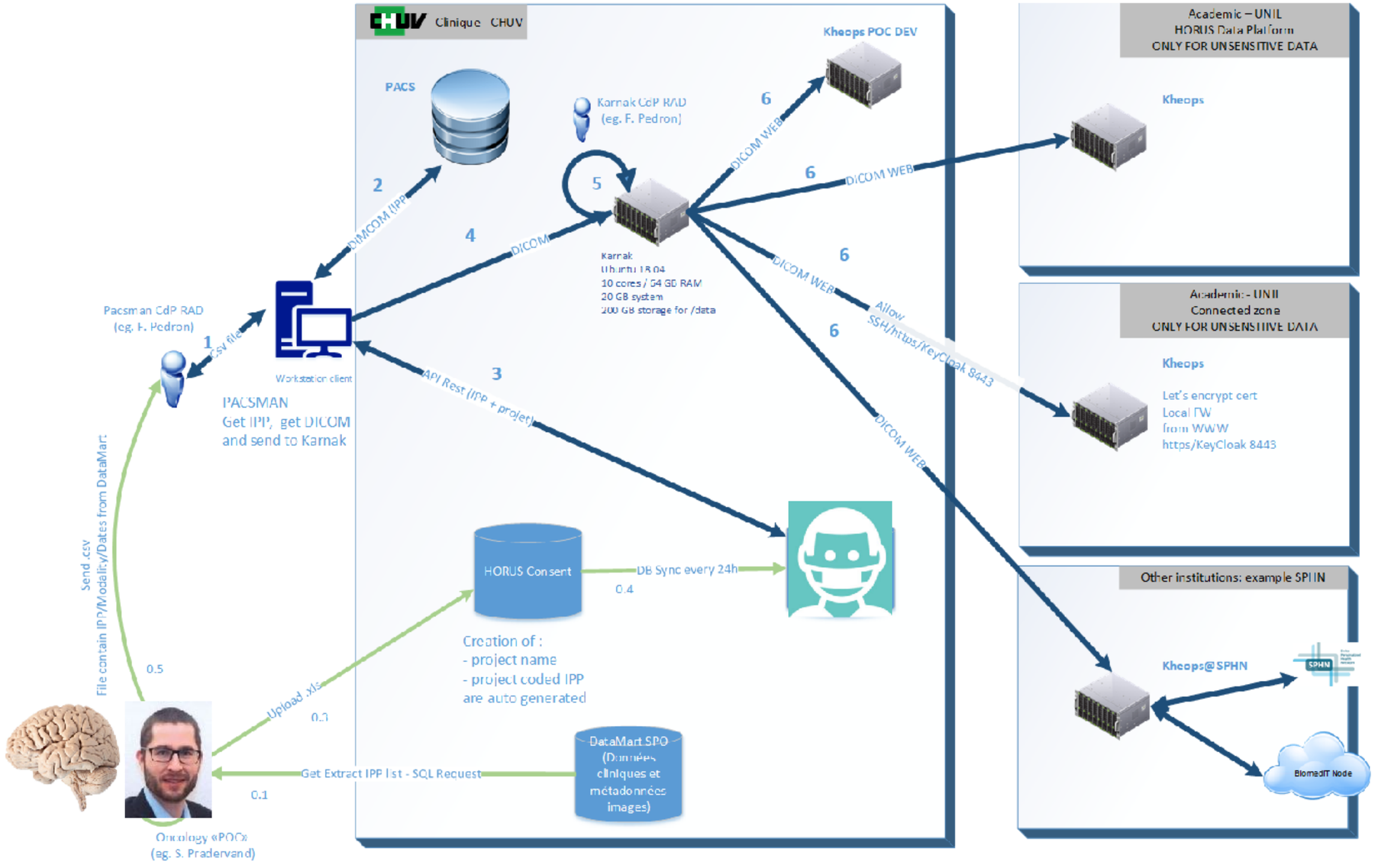
HORUS Image


Clinique CHUV
PACS
Karnak CdP RAD
(eg. F. Pedron)
Kheops POC DEV
Karnak
Ubuntu 18.04
10 cores / 64 GB RAM
20 GB system
200 GB storage for /data
DICOM (IPP)
DICOM
DICOM WEB
API Rest (IPP + project)
Pacsman CdP RAD
(eg. F. Pedron)
csv file
Workstation client
PACSMAN
Get IPP, get DICOM
and send to Karnak
Send .csv
File contain IPP/Modality/Dates from DataMart
HORUS Consent
DB Sync every 24h
HORUS
warehouse
(datawarehouse
recherche &
Graph)
Creation of :
- project name
- project coded IPP
are auto generated
Upload .xls
Export
Get Extract IPP list - SQL Request
DataMart SPO
(Données
cliniques et
métadonnées
images)
Oncology «POC»
(eg. S. Pradervand)
Academic – UNIL
HORUS Data Platform
ONLY FOR UNSENSITIVE DATA
Kheops
Academic - UNIL
Connected zone
ONLY FOR UNSENSITIVE DATA
Kheops
Let's encrypt cert
Local FW
from WWW
https/KeyCloak 8443
Allow
SSH/https/KeyCloak 8443
Other institutions: example SPHN
Kheops@SPHN
SPHN
BiomedIT Node

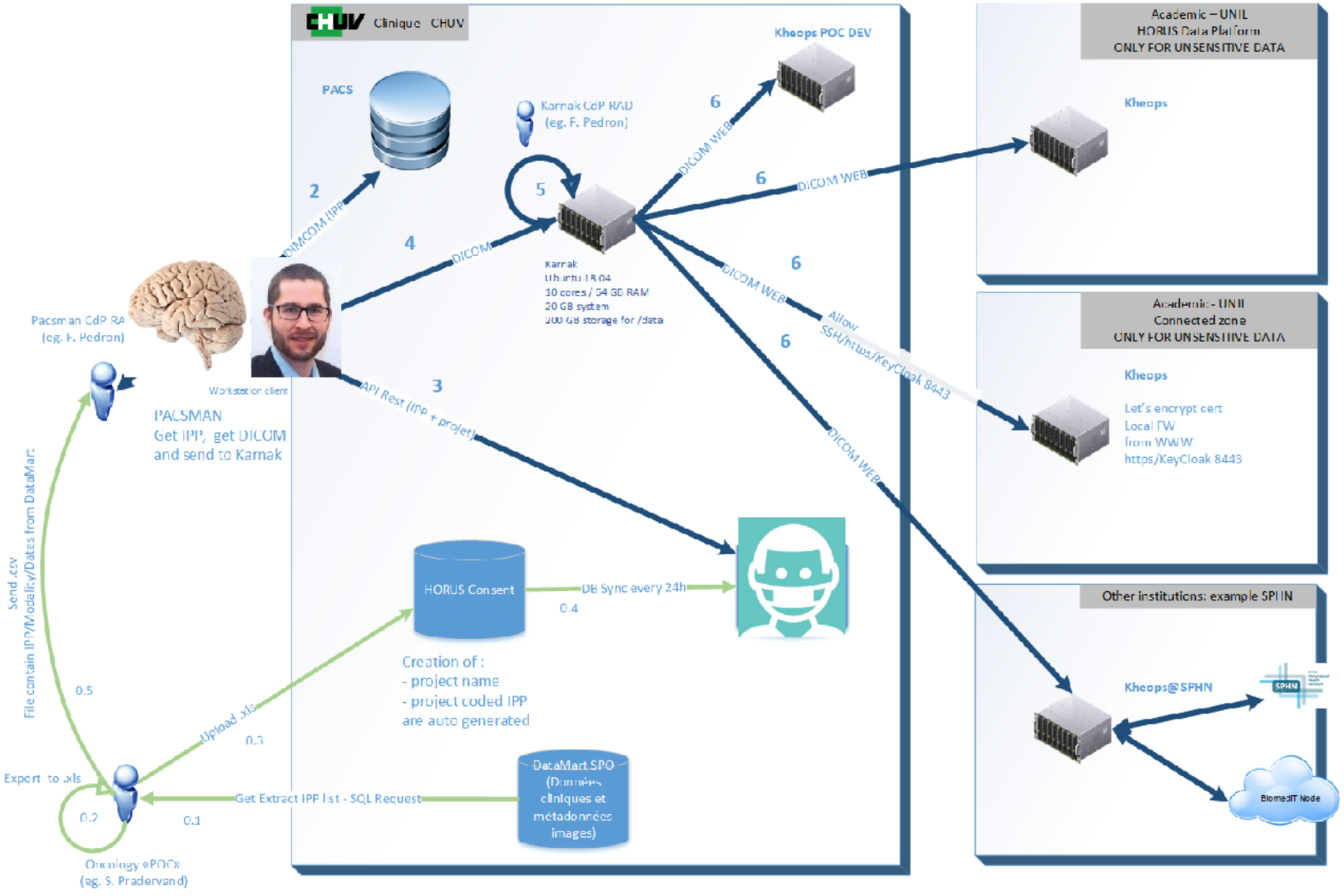
HORUS Image


Clinique CHUV
PACS
Karnak CdP RAD
(eg. F. Pedron)
Kheops POC DEV
Karnak
Ubuntu 18.04
10 cores / 64 GB RAM
20 GB system
200 GB storage for /data
DICOM (IPP)
DICOM
DICOM WEB
Allow SSH/https/KeyCloak 8443
API Rest (IPP + project)
Pacsman CdP RAD
(eg. F. Pedron)
csv file
Workstation client
PACSMAN
Get IPP, get DICOM
and send to Karnak
Send .csv
File contain IPP/Modality/Dates from DataMart
Creation of :
- project name
- project coded IPP
are auto generated
DB Sync every 24h
HORUS
warehouse
(datawarehouse
recherche &
Graph)
Upload .xls
Export
Get Extract IPP list - SQL Request
DataMart SPO
(Données
cliniques et
métadonnées
images)
Oncology «POC»
(eg. S. Pradervand)
Academic – UNIL
HORUS Data Platform
ONLY FOR UNSENSITIVE DATA
Kheops
Academic - UNIL
Connected zone
ONLY FOR UNSENSITIVE DATA
Kheops
Let's encrypt cert
Local FW
from WWW
https/KeyCloak 8443
Other institutions: example SPHN
Kheops@SPHN
SPHN
BiomedIT Node

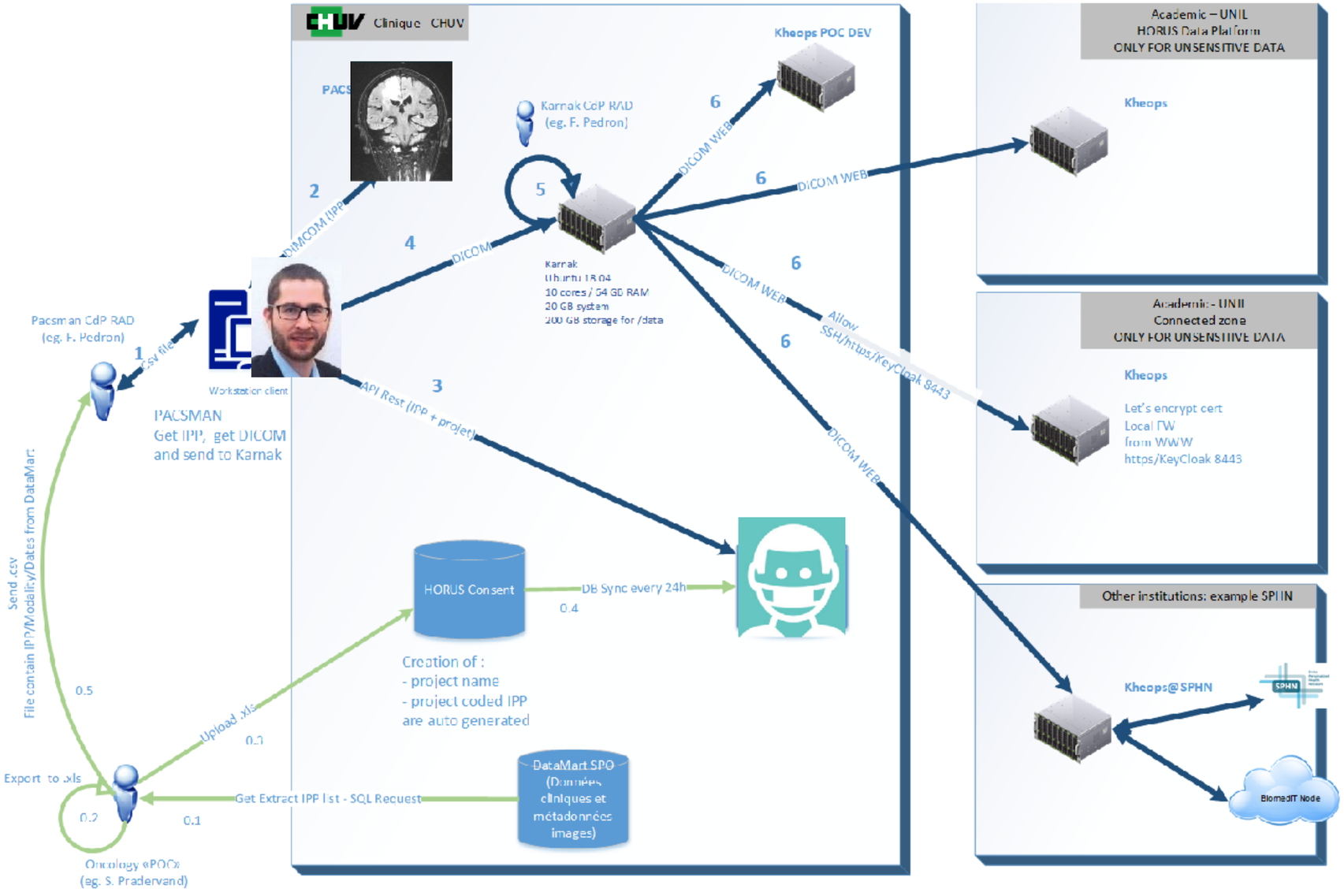
HORUS Image


Clinique CHUV
PACS
Karnak CdP RAD
(eg. F. Pedron)
Kheops POC DEV
Academic – UNIL
HORUS Data Platform
ONLY FOR UNSENSITIVE DATA
Kheops
Academic - UNIL
Connected zone
ONLY FOR UNSENSITIVE DATA
Kheops
Let's encrypt cert
Local FW
from WWW
https/KeyCloak 8443
Other institutions: example SPHN
Kheops@SPHN
BiomedIT Node
Pacsman CdP RAD
(eg. F. Pedron)
Workstation client
PACSMAN
Get IPP, get DICOM
and send to Karnak
Karnak
Ubuntu 18.04
10 cores / 64 GB RAM
20 GB system
200 GB storage for /data
DICOM (IPP)
DICOM
DICOM WEB
Allow SSH/https/KeyCloak 8443
API Rest (IPP + project)
HORUS Consent
DB Sync every 24h
Creation of :
- project name
- project coded IPP
are auto generated
DataMart SPO
(Données cliniques et métadonnées images)
Get Extract IPP list - SQL Request
Upload .xls
Send .csv
File contain IPP/Modality/Dates from DataMart
Oncology «POC»
(eg. S. Pradervand)

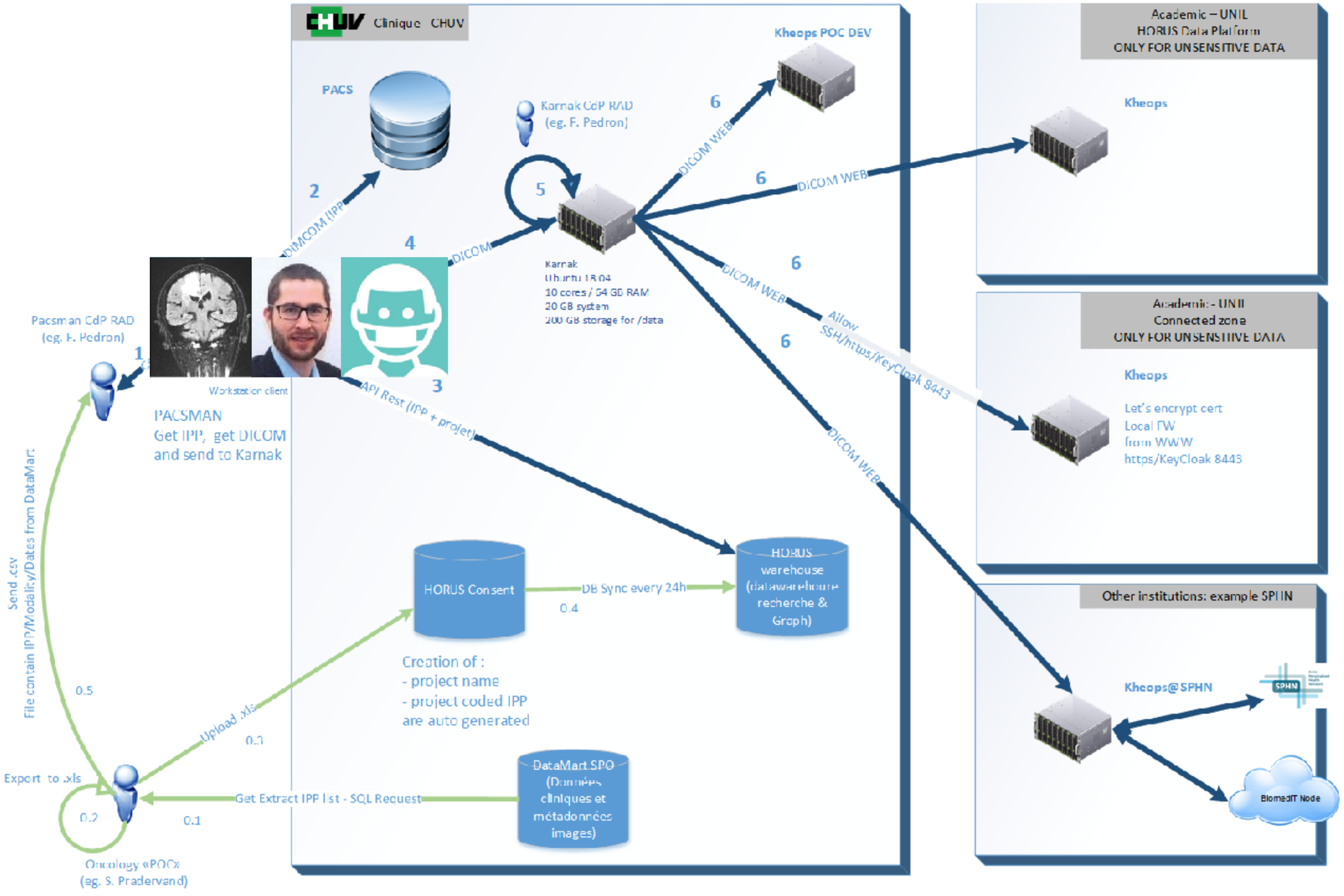
HORUS Image


Clinique CHUV
PACS
Karnak CdP RAD
(eg. F. Pedron)
Kheops POC DEV
Karnak
Ubuntu 18.04
10 cores / 64 GB RAM
20 GB system
200 GB storage for /data
DICOM (IPP)
DICOM
DICOM WEB
Allow SSH/https/KeyCloak 8443
API Rest (IPP + project)
Pacsman CdP RA
(eg. F. Pedron)
Workstation client
PACSMAN
Get IPP, get DICOM
and send to Karnak
HORUS Consent
DB Sync every 24h
Creation of :
- project name
- project coded IPP
are auto generated
Send .csv
File contain IPP/Modality/Dates from DataMart
Upload .xls
Export to .xls
Get Extract IPP list - SQL Request
DataMart SPO
(Données cliniques et métadonnées images)
Oncology «POC»
(eg. S. Pradervand)
Academic – UNIL
HORUS Data Platform
ONLY FOR UNSENSITIVE DATA
Kheops
Academic - UNIL
Connected zone
ONLY FOR UNSENSITIVE DATA
Kheops
Let's encrypt cert
Local FW
from WWW
https/KeyCloak 8443
Other institutions: example SPHN
Kheops@SPHN
SPHN
BiomedIT Node

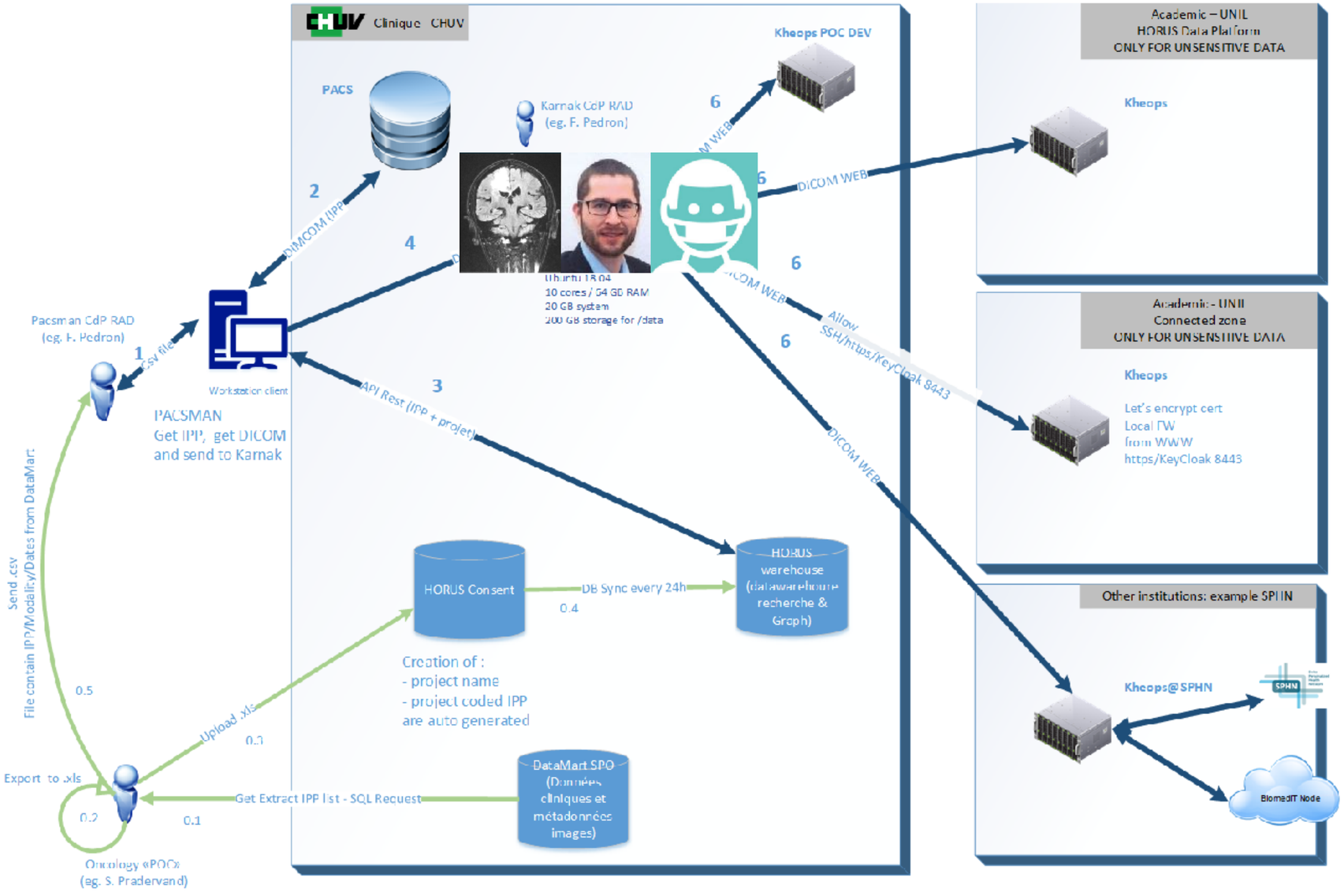
HORUS Image


Clinique CHUV
Kheops POC DEV
Karnak CdP RAD
(eg. F. Pedron)
Karnak
Ubuntu 18.04
10 cores / 64 GB RAM
20 GB system
200 GB storage for /data
Pacsman CdP RAD
(eg. F. Pedron)
Workstation client
PACSMAN
Get IPP, get DICOM
and send to Karnak
API Rest (IPP + project)
DICOM
DICOM WEB
Allow SSH/https/KeyCloak 8443
HORUS Consent
DB Sync every 24h
Creation of :
- project name
- project coded IPP
are auto generated
DataMart SPO
(Données cliniques et métadonnées images)
Get Extract IPP list - SQL Request
Upload .xls
Export to .xls
Oncology «POC»
(eg. S. Pradervand)
Send .csv
File contain IPP/Modality/Dates from DataMart
Academic – UNIL
HORUS Data Platform
ONLY FOR UNSENSITIVE DATA
Kheops
Academic - UNIL
Connected zone
ONLY FOR UNSENSITIVE DATA
Kheops
Let's encrypt cert
Local FW
from WWW
https/KeyCloak 8443
Other institutions: example SPHN
Kheops@SPHN
SPHN
BiomedIT Node

HORUS Image


Clinique CHUV
PACS
Karnak CdP RAD
(eg. F. Pedron)
Kheops POC DEV
Karnak
Ubuntu 18.04
10 cores / 64 GB RAM
20 GB system
200 GB storage for /data
DICOM (IPP)
DICOM
DICOM WEB
Allow SSH/https/KeyCloak 8443
API Rest (IPP + project)
Pacsman CdP RAD
(eg. F. Pedron)
Workstation client
PACSMAN
Get IPP, get DICOM
and send to Karnak
Send .csv
File contain IPP/Modality/Dates from DataMart
HORUS Consent
DB Sync every 24h
HORUS warehouse (datawarehouse recherche & Graph)
Creation of :
- project name
- project coded IPP
are auto generated
Upload .xls
Export to .xls
Get Extract IPP list - SQL Request
DataMart SPO (Données cliniques et métadonnées images)
Oncology «POC»
(eg. S. Pradervand)
Academic – UNIL
HORUS Data Platform
ONLY FOR UNSENSITIVE DATA
Kheops
Academic - UNIL
Connected zone
ONLY FOR UNSENSITIVE DATA
Kheops
Let's encrypt cert
Local FW
from WWW
https/KeyCloak 8443
Other institutions: example SPHN
Kheops@SPHN
SPHN
BiomedIT Node

HORUS Image


Clinique CHUV
Kheops POC DEV
PACS
Karnak CdP RAD
(eg. F. Pedron)
6
DICOM WEB
2
DIMCOM (IPP
4
Ubuntu 18.04
10 cores / 64 GB RAM
20 GB system
200 GB storage for /data
Pacsman CdP RAD
(eg. F. Pedron)
1
csv file
Workstation client
PACSMAN
Get IPP, get DICOM
and send to Karnak
3
API Rest (IPP + project)
Allow
SSH/https/KeyCloak 8443
Send .csv
File contain IPP/Modality/Dates from DataMart
0.5
HORUS Consent
DB Sync every 24h
0.4
HORUS
warehouse
(datawarehouse
recherche &
Graph)
Creation of :
- project name
- project coded IPP
are auto generated
Upload .xls
0.3
Export to .xls
0.2
0.1
Get Extract IPP list - SQL Request
DataMart SPO
(Données
cliniques et
métadonnées
images)
Oncology «POC»
(eg. S. Pradervand)
Academic – UNIL
HORUS Data Platform
ONLY FOR UNSENSITIVE DATA
Kheops
Academic - UNIL
Connected zone
ONLY FOR UNSENSITIVE DATA
Kheops
Let's encrypt cert
Local FW
from WWW
https/KeyCloak 8443
Other institutions: example SPHN
Kheops@SPHN
SPHN
BiomedIT Node

HORUS Image

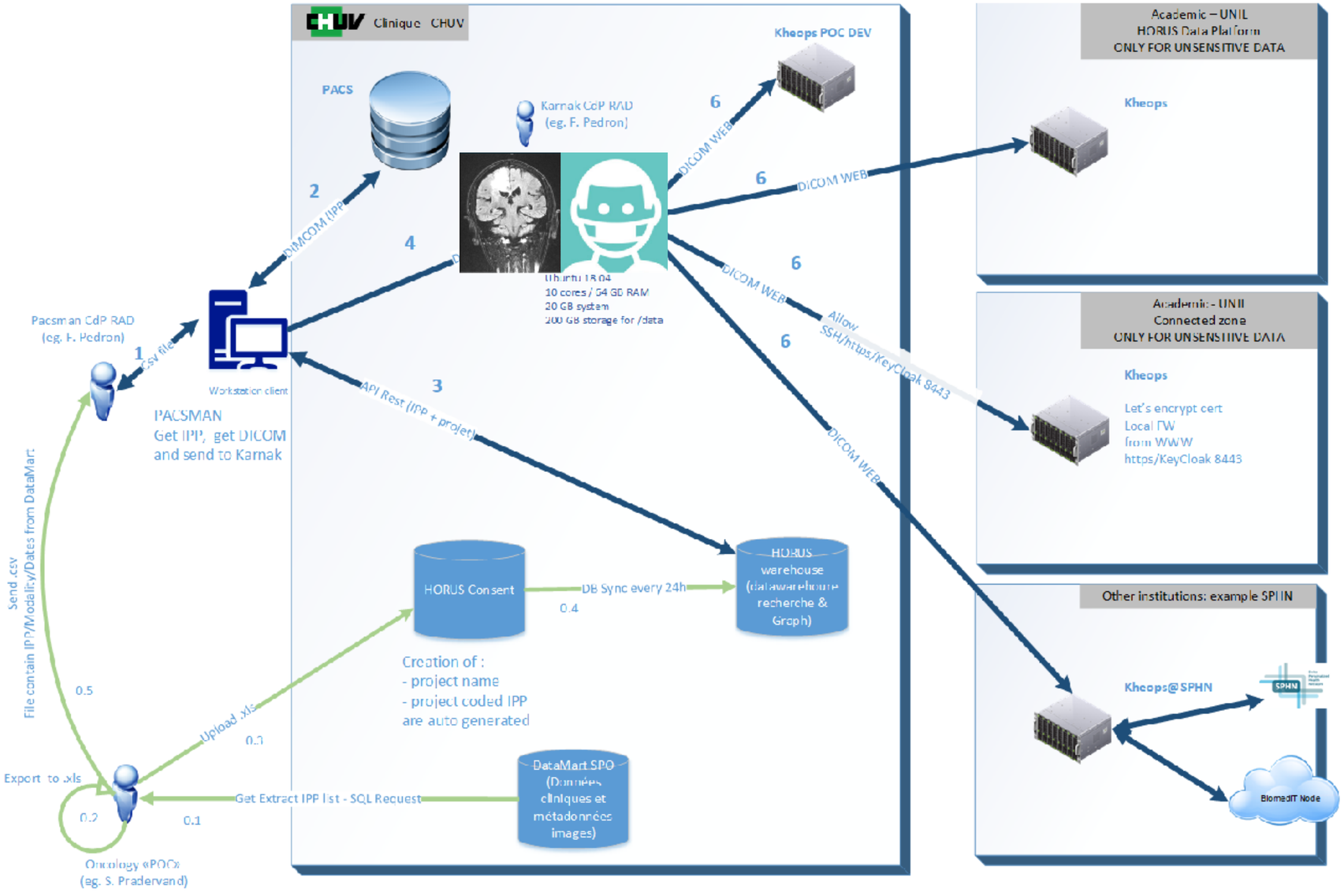
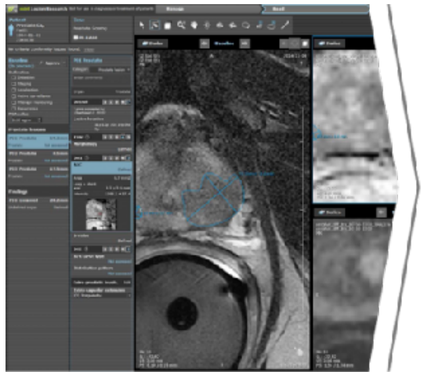

Clinique CHUV
PACS
Karnak CdP RAD
(eg. F. Pedron)
Kheops POC DEV
Ubuntu 18.04
10 cores / 64 GB RAM
20 GB system
200 GB storage for /data
Pacsman CdP RAD
(eg. F. Pedron)
Workstation client
PACSMAN
Get IPP, get DICOM
and send to Karnak
DICOM (IPP)
DICOM WEB
API Rest (IPP + project)
Allow SSH/https/KeyCloak 8443
HORUS Consent
DB Sync every 24h
HORUS warehouse (datawarehouse recherche & Graph)
Creation of :
- project name
- project coded IPP
are auto generated
DataMart SPO (Données cliniques et métadonnées images)
Get Extract IPP list - SQL Request
Upload .xls
Export to .xls
Oncology «POC»
(eg. S. Pradervand)
Send .csv
File contain IPP/Modality/Dates from DataMart
Academic – UNIL
HORUS Data Platform
ONLY FOR UNSENSITIVE DATA
Kheops
Academic - UNIL
Connected zone
ONLY FOR UNSENSITIVE DATA
Kheops
Let's encrypt cert
Local FW
from WWW
https/KeyCloak 8443
Other institutions: example SPHN
Kheops@SPHN
SPHN
BiomedIT Node

Image annotation

Label the images?

I need a manually traced contour of liver cancer tumors

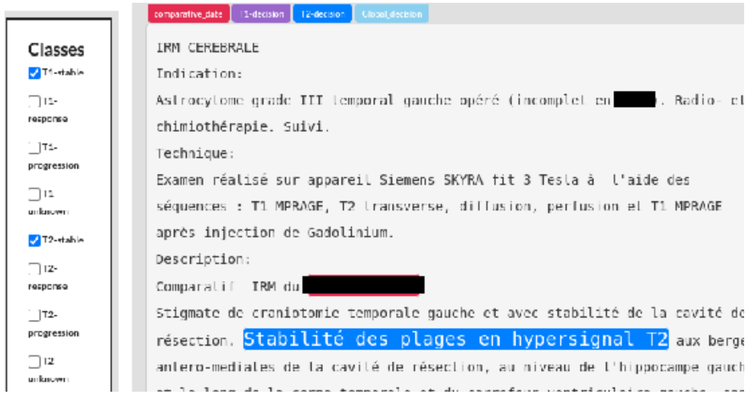


[Mint medical]

weak labels

Label the reports?

I need a document-level label and some entities tagged to determine glioma response or progression



[Di Noto et al, PharML 2021]

Collaborations with Naïk Vietti Violi, Clarisse Dromain, Meri Bach Cuadra, Patric Hagmann, Chirine Attat, Alexandre Teiga, Andreas Hottinger, Monika Hegi

Model development

ML = Theory + Data + Labels + Code + Compute
✓ ✓

Learning the basics

- *Hands-on intro to ML for biomedical data* - 1.5 ECTS doctoral school course, 2x/year
- *Introduction to ML for biologists* - 1 ECTS BSc/MSc course, 1x/year
- *Gentle Introduction to Decision Tree and Random Forest with Python and R*, DCSR course, periodic
- *A Gentle Introduction to Deep Learning with Python and R*, DCSR course, periodic

Keeping up with state of the art

- *Deep learning journal club* - 1/month
- *ML Café* - quarterly

12 general nodes

RAM: 128 GB
CPU: 2x E5-2650v2, 8 cores

2 GPU nodes

RAM: 1TB
CPU: 2x Xeon Gold 6148, 40 cores
GPU: 4x RTX 2080, 1x V100

RAM: 2TB
CPU: 2x EPYC 7742, 128 cores
GPU: 2x RTX 3090

CHUV HPC

Jura - 10 nodes

RAM: 256 GB - 1 TB
cores: 32-160
Xeon Phi: 4 nodes

Curnagl - 72 nodes

RAM: 512 GB - 1 TB
CPU: 2x EPYC 7402, 24 cores
GPU: 8x2 V100

UNIL HPC

ML workloads for imaging

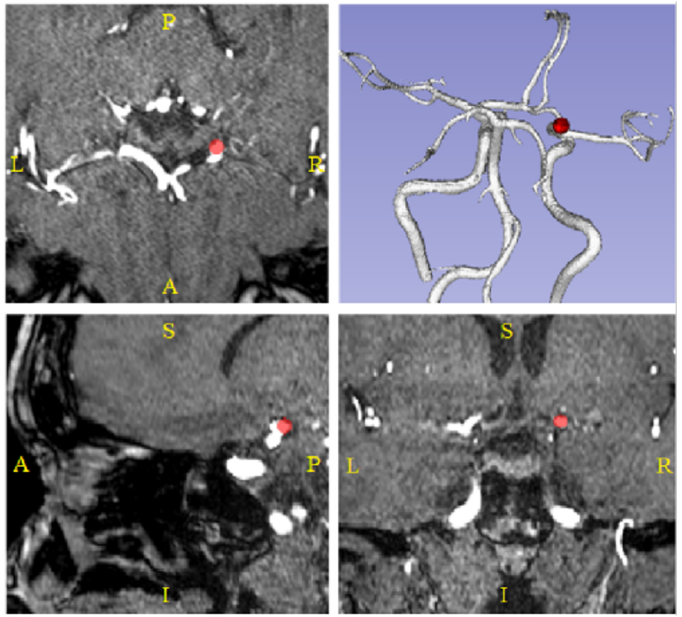
Generic pipeline

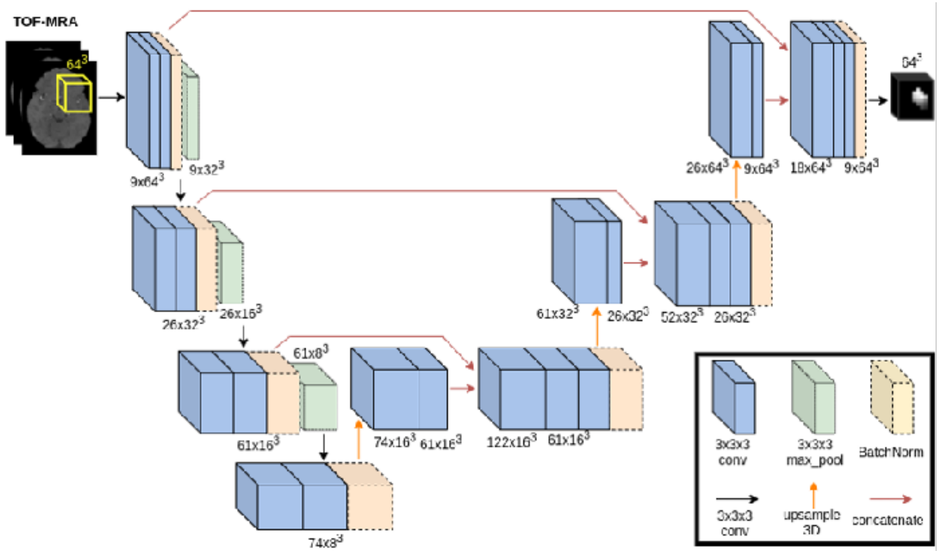
Data prep

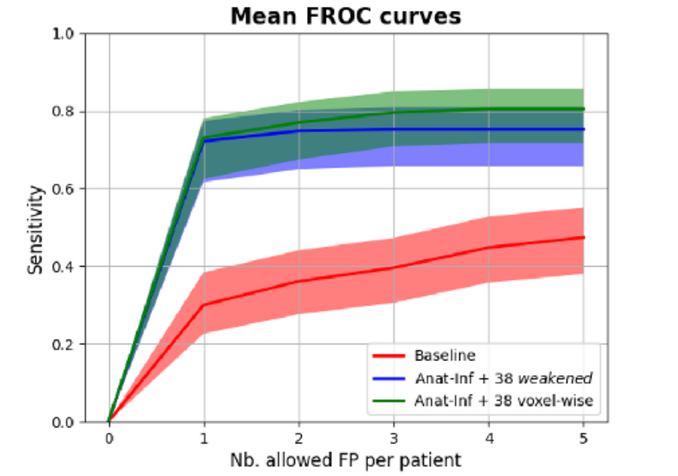
ML

Workflow engines

GPU-bound: Aneurysm detection







Task: locate aneurysms (N=198)

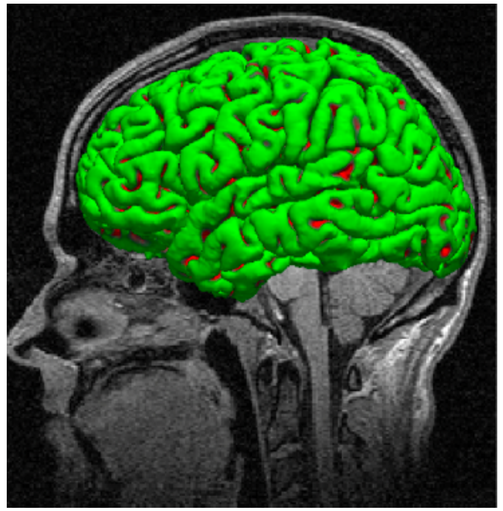
Approach: detection-by-segmentation

Model: 855K params U-Net, **training time 14h x 5** + tuning (1x RTX 3090 48 GB)

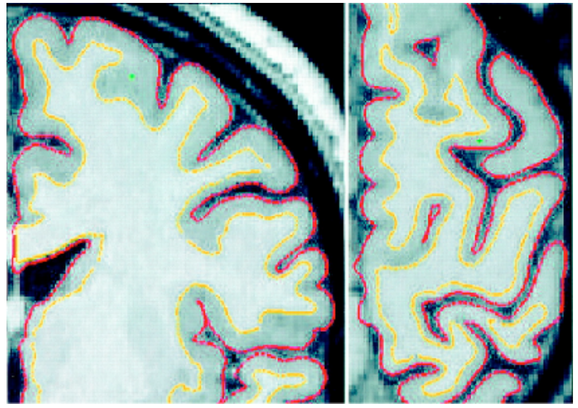
Results: 80% sens @ 1.2 FP (in house), 68% @ 2.5 FP (challenge, rank 4/18)

[Di Noto et al, MCLN 2020] [Di Noto et al., submitted (arXiv:2103.06168)]

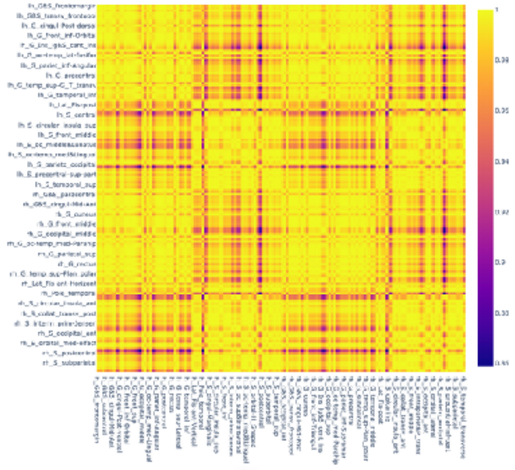
CPU-bound: Brain morphometry for diagnosis



[Sandu et al., 2014]



[Fischl & Dale, 2001]



[Mahdi, 2020]

93% acc
Alzheimer
vs controls

Task: segment the brain
Approach: FreeSurfer
Runtime: **12 CPU hours** (Xeon Gold) per subject - but **embarassingly parallel**

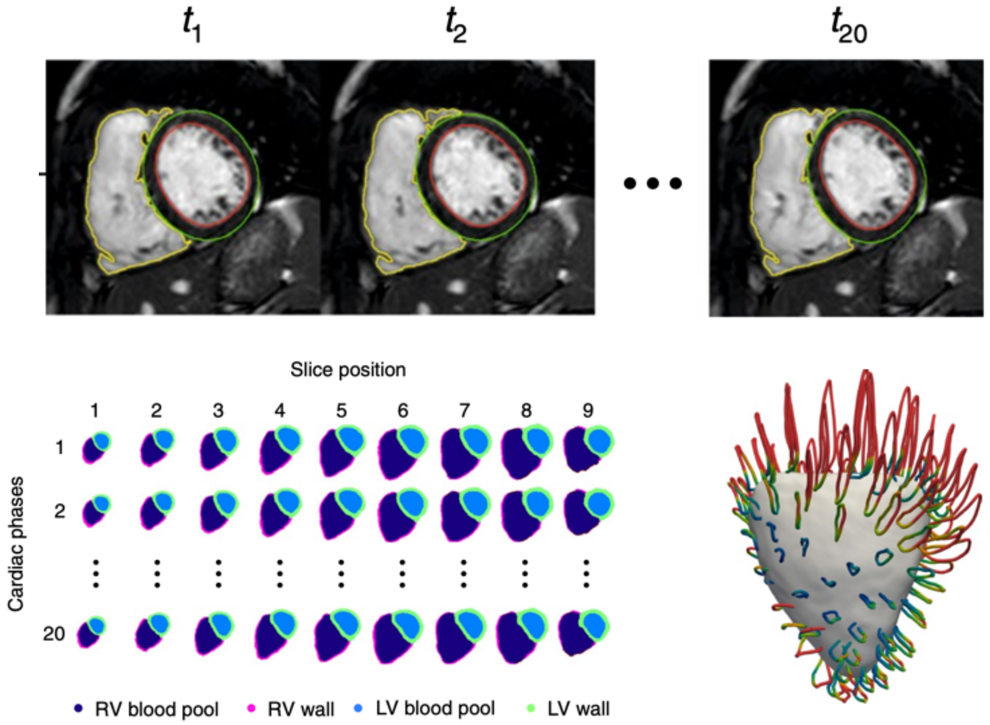
graph-based representation: morphometric similarity network

ML: graph embedding + non-linear dimensionality reduction + SVM

SNSF project partnership - PI J. Popp

“storage-bound”: Cardiac MRI processing for imaging genetics

[Bello et al 2019]



Task: segment the heart & estimate motion

Unprocessed data: ~ 2.5 GB / subject

x 42’000…

x 2 (processed)…

~ 210 TB

SNSF project CRSII5_202276, collab. with R. van Heeswijk, R. Hullin, P. Meyer, M. Stuber, Z. Kutalik, M. Cuendet, R. Gatta, A. Thomas, J.P. Vallée, D. Carballo

Wrap-up and take-home points

using clinical routine images for ML…

requires tricky depersonalisation to implement due to minute details (and manufacturer differences) in DICOMs

needs considerably more ethical care than typical computer vision problems

is increasingly feasible and scalable thanks to open software and standardisation efforts

unlocks literally millions of images for **very impactful research,** both basic and applied*

*See related workshops and conferences
https://mlcnws.com/
https://sites.google.com/view/pharml2021
https://www.miccai2021.org
https://www.midl.io/
…

Thanks

Translational Machine Learning Lab

Veronica Ravano, Dr Jonathan Patiño-Lopez, Dr Jaume Banus Cobo, Dr Elda Fischi-Gomez, Xavier Sieber, Tommaso Di Noto (+ Costa Georgantas)

Collaborators

Solange Zoergiebel
Alexandre Wetzel
Patrick Zosso
Frédéric Pedron
Nicolas Rosat
Firas Ben Othman
Roberto Fabbretti + team

Funding

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Swiss Confederation

Innosuisse – Swiss Innovation Agency

FNSNF