Infrastructure 'Information Analysis and Management' (IAM)

Christian BARILLOT
FLI-IAM Chair (Rennes)

Michel DOJAT
FLI-IAM co-Chair (Grenoble)

Michael KAIN
FLI-IAM Technical Manager (Inria)

Many Good Reasons for Sharing Data and Tools in *In Vivo* Imaging

At Least 3 ….

- **Scientific**

« Power failure: why small sample size undermines the reliability of neuroscience »
Button et al. Nat. Neurosc 2013
« Why most published research findings are false" **Ioannidis Plos Med 2005**
"Data from preclinical animal studies appear to be associated with even greater bias »
Ioannidis ScienceTranslationalMed 2012,
« Why small low-powered studies are worse than large high-powered studies and how to protect against trivial findings in research **» Ingre Neuroimage 2012**

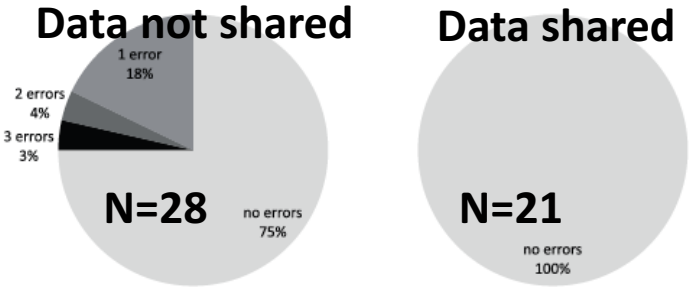
- **Underpowered is endemic** (false positive, false negative)

Many Good Reasons for Sharing Data and Tools in *In Vivo* Imaging

At Least 3 ….

- Scientific
  - **Data analysis and reporting are often selective and biased**

Errors in reporting statistical results



Wicherts et al. PLOS One 2011

Many Good Reasons for Sharing Data and Tools in *In Vivo* Imaging

At Least 3 ….

- Scientific
  - **Rarely replicated so false solutions persist**
  - **New questions, increasing the speed of discovery**

Large cohort, Re-analysis, Replication, MetaAnalysis

« collective mind » Fox et al. Ann Rev Neurosci 2014

« crowd science » Franzoni & Suerman Res Policy 2014

Many Good Reasons for Sharing Data and Tools in *In Vivo* Imaging

- **Economic**

Data Acquisition: e.g. neuroimaging 300€/H 30 subjects x 2 groups 18 K€ …
Many efforts for image processing tools development and validation
Avoid reinventing existing data and tools ….

Reducing cost doing science, maximize investment

Many Good Reasons for Sharing Data and Tools in *In Vivo* Imaging

- **Economic**

Data Acquisition: e.g. neuroimaging 300€/H 30 subjects x 2 groups 18 K€ …
Many efforts for image processing tools development and validation
dont replicate existing data and tools ….

→ **Reducing cost doing science, maximize investment**

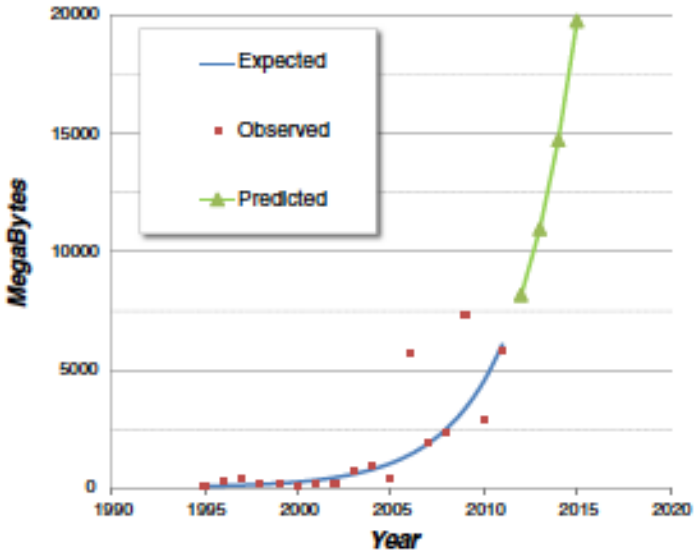
- **Ethic**

Healthy subject and patients give their consent for contributing to knowledge advancement.
For preclinical studies, ethical obligation to ensure that the animals are not wasted.

→ **Maximize their participation**

NeuroImage, indicates that since 1995 the amount of data collected has doubled approximately every 26 months

This is likely to be an under-estimate for raw dataset sizes: advances in MRI physics are accelerating the pace at which data can be aquired per unit time.

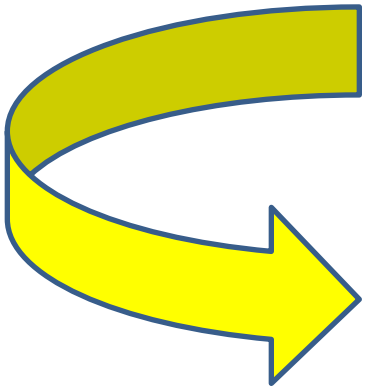


Van Horn & Toga Brain Im Behav 2014

(1) respect for participant privacy and confidentiality;
(2) logistics of data accessibility;
And
(3) long-term sustainability.

Researchers, journals and funding agencies must work together to identify and implement technical solutions that allow the most effective data sharing **without greatly increasing the burden on researchers.**

Poldrack and Gorgolewski Nat Neuro 2014



FLI-IAM

France Life Imaging (FLI)

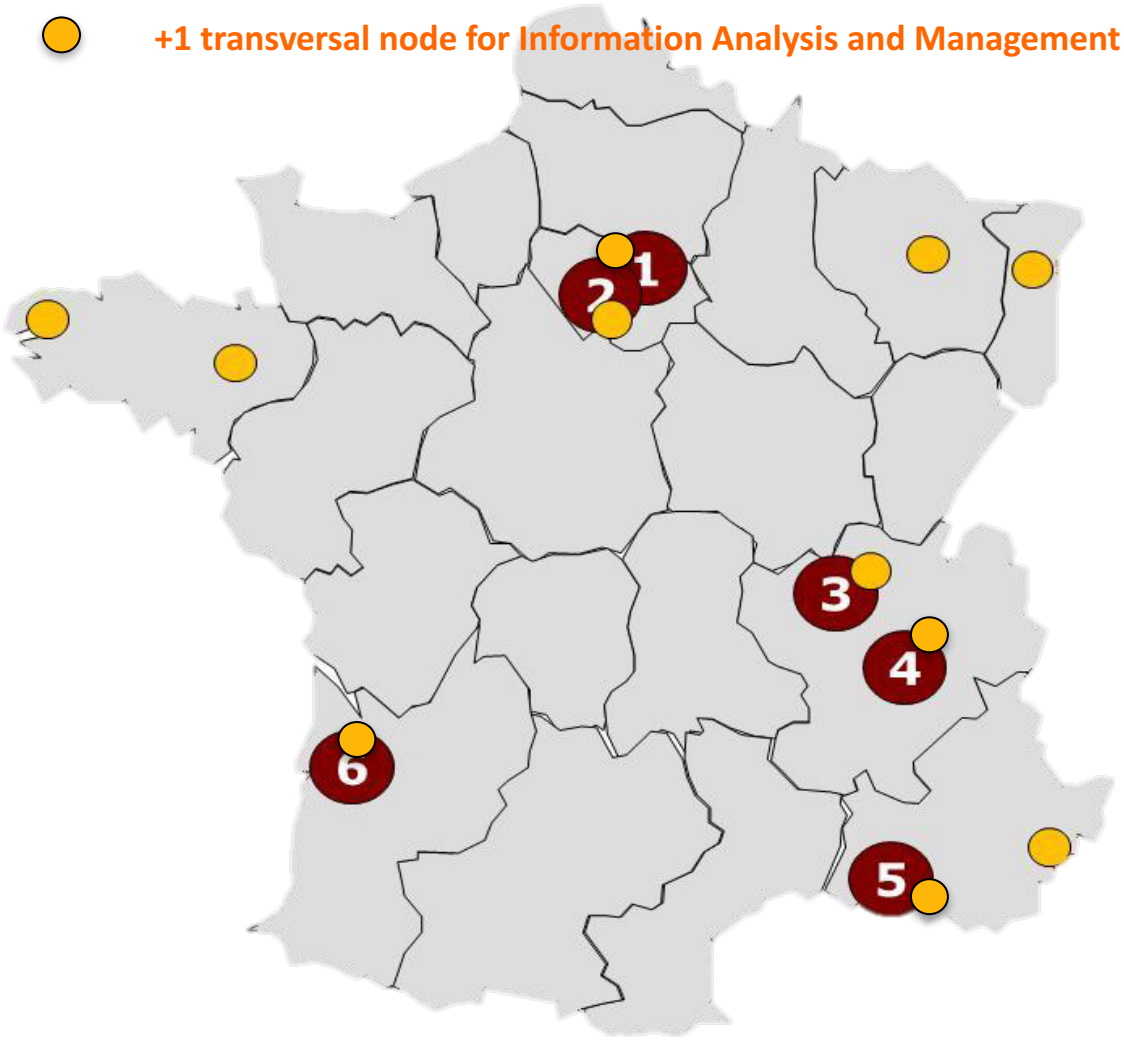
- **France Life Imaging (FLI) is a large-scale research infrastructure project, aimed at establishing a coordinated and harmonized network of biomedical imaging in France: http://francelifeimaging.fr**
- **Its mission is to:**
  - **coordinate** nation-wide research activities concerned with in vivo imaging and combine the skills to push the current technological barriers
  - provide scientists a **convenient access to a complete range of imaging technologies** (150 imaging systems) and integrated services; in addition, the infrastructure will be open to collaborations with industrial partners
  - FLI will also propose training opportunities associated with imaging platforms. FLI complements **France BioImaging** for *in-vitro* Imaging (FBI).

- **2012, Investissements d'Avenir - Infrastructure en Biologie et Santé**
- **37 million Euros for 8 years**

France Life Imaging

6 physical nodes

+1 transversal node for Information Analysis and Management



Information Analysis and Management (IAM) - Goals

- **Objectives of the node:**
  - Specify and set up a hardware and software infrastructure for the management and processing of *in vivo* images
  - Elaborate and implement usage scenarios of the infrastructure
  - Set-up an operational structure to operate the infrastructure, that can be self-funded from internal seed funding

- **For:**
  - **Clinician**: to conduct large and/or clinical and preclinical research studies, involving new innovative in-vivo medical imaging and therapeutic procedures
  - **Pharma**: to provide pharma and CRO companies high technological computational solutions for *in vivo* imaging
  - **Medical imaging community**: to allow experimentation and validation of new innovative in-vivo imaging solutions

- Using strengths: based on **existing** high technological expertise and experience:
  - From data management solutions: ArchiMed, CATI-DB, SHANOIR …
  - From medical image processing solutions: BrainVisa, medInria, VIP, …
  - In cooperation with large national clinical cohorts: CATI, OFSEP, …

Information Analysis and Management (IAM) - Goals

- **Objectives of the node:**
  - Specify and set up a hardware and software infrastructure for the management and processing of *in vivo* images
  - Elaborate and implement usage scenarios of the infrastructure
  - Set-up an operational structure to operate the infrastructure, that can be self-funded from internal seed funding

- **For:**
  - **Clinician**: to conduct large and/or clinical and preclinical research studies, involving new innovative in-vivo medical imaging and therapeutic procedures
  - **Pharma**: to provide pharma and CRO companies high technological computational solutions for *in vivo* imaging
  - **Medical imaging community**: to allow experimentation and validation of new innovative in-vivo imaging solutions

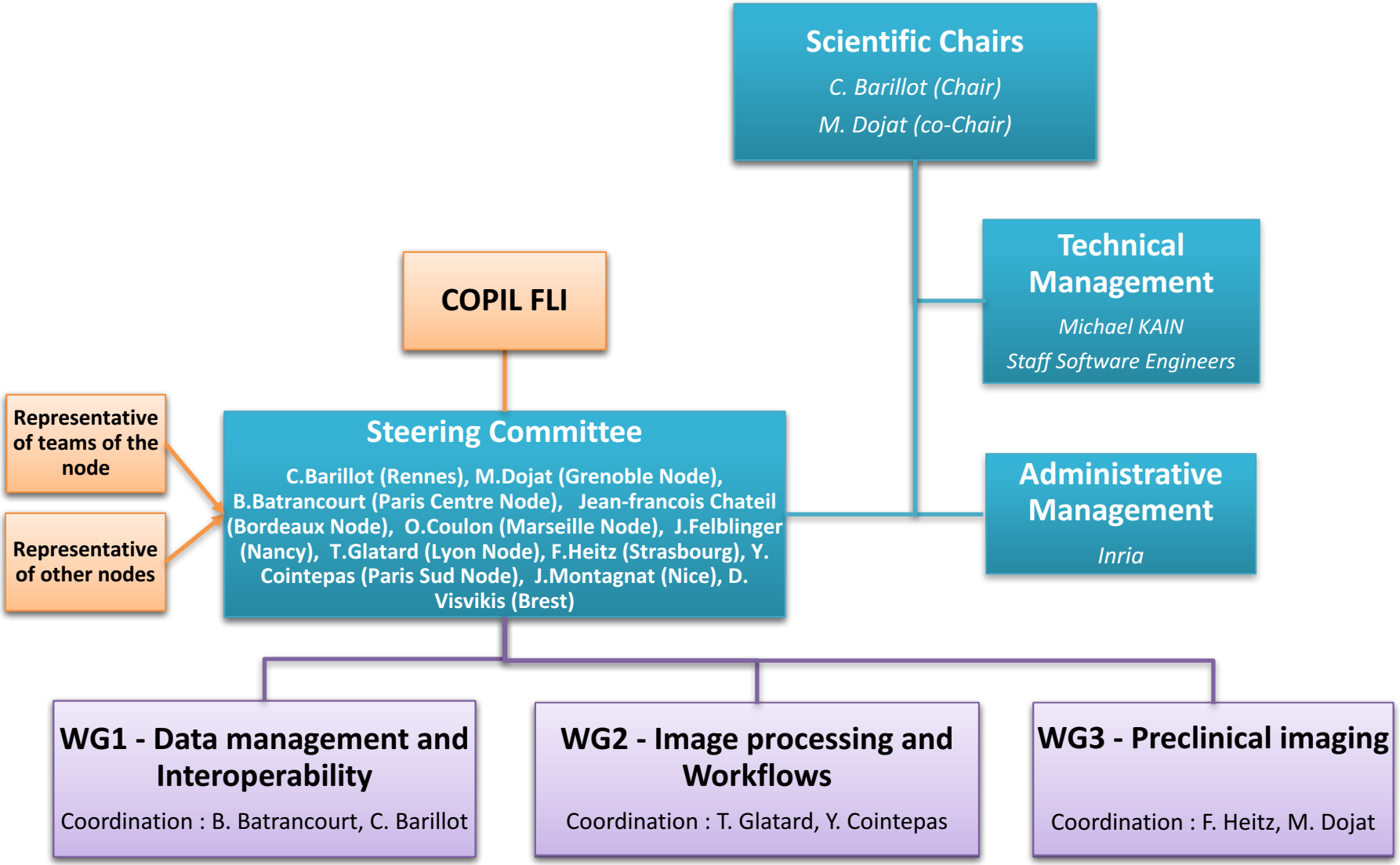
- Using strengths: based on **existing** high technological expertise and experience:
  - From data management solutions: ArchiMed, CATI-DB, SHANOIR …
  - From medical image processing solutions: BrainVisa, medInria, VIP, …
  - In cooperation with large national clinical cohorts: CATI, OFSEP, …

The IAM node consists in developing a versatile software platform made up of several sub-components, which allows inter-connection of hardware facilities and software platforms.

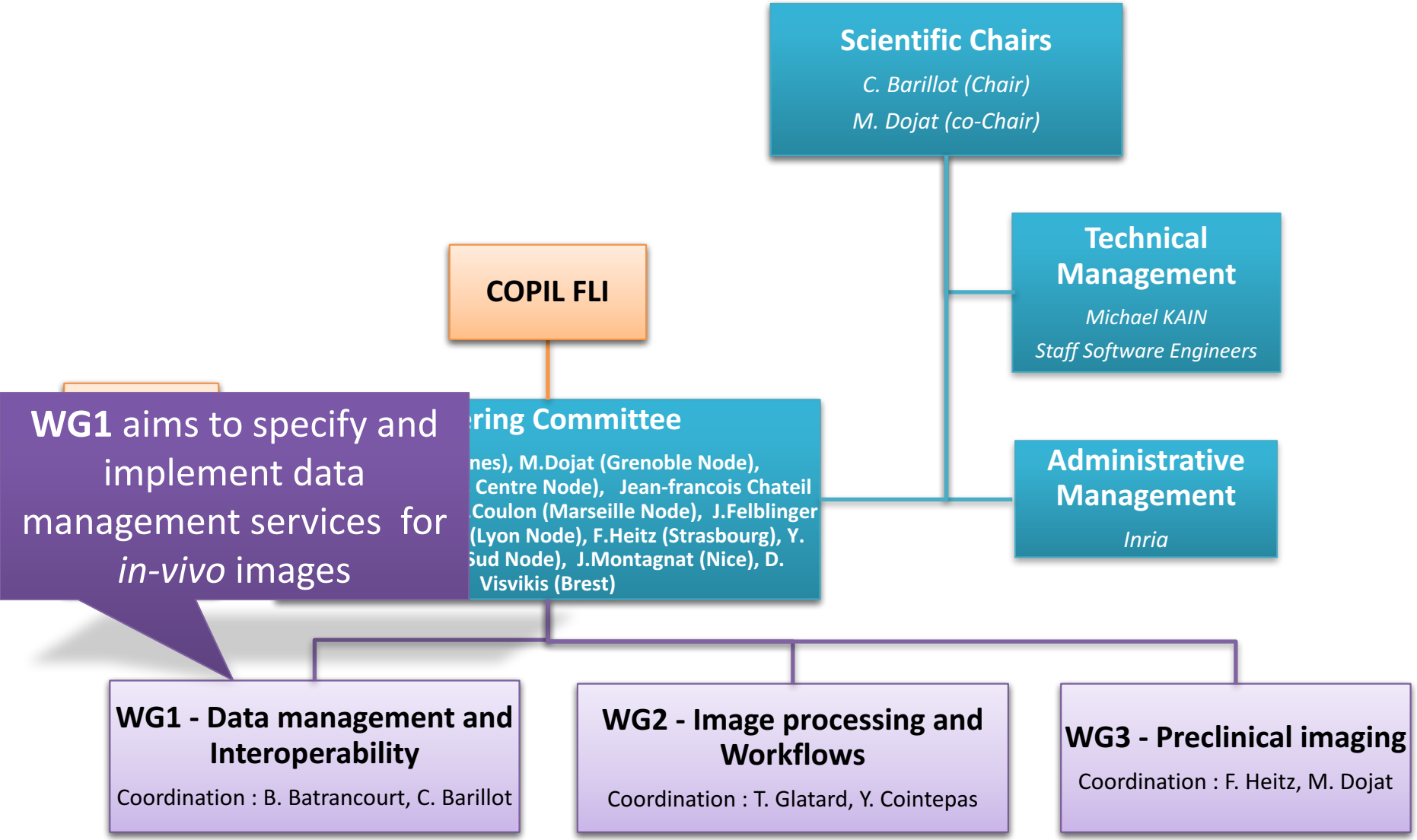
Information Analysis and Management (IAM) - Facts

- **Founding partners:**
  - Top national teams in medical imaging:
    CEA (Paris), **CNRS** (Bordeaux, Lyon, Strasbourg, Marseille), **INRIA** (Rennes), **INSERM** (Brest, Grenoble, Nancy, Paris)
- **Currently 15 engineers working full-time, in 8 cities all over France**
- **Deployment stages:**
  - Initial stage (2013 – end of 2017/18), **phase 1**:
    - Set up of the infrastructure, specification of the needs, roadmap definition for the development and the exploitation of the infrastructure based on the roadmap
    - Start a pilot phase on a limited number of clinical and preclinical research application domains
  - Operational stage (after 2017/18), **phase 2**:
    - **Open call** to select industrial management and exploitation providers of the developed infrastructure (including SME)
    - Extend the infrastructure to broaden the range of clinical and preclinical research application domains

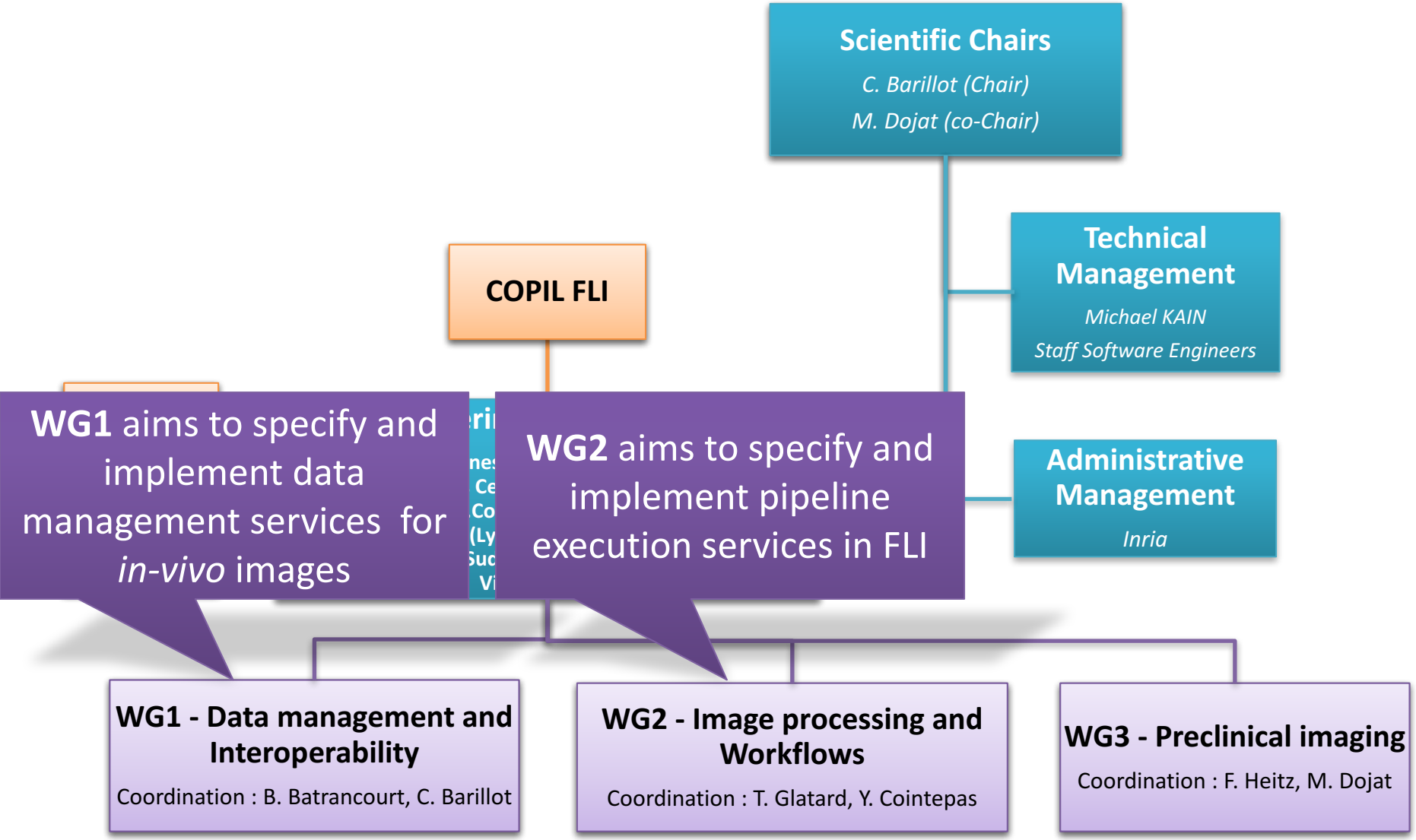
FLI-IAM Node Organization



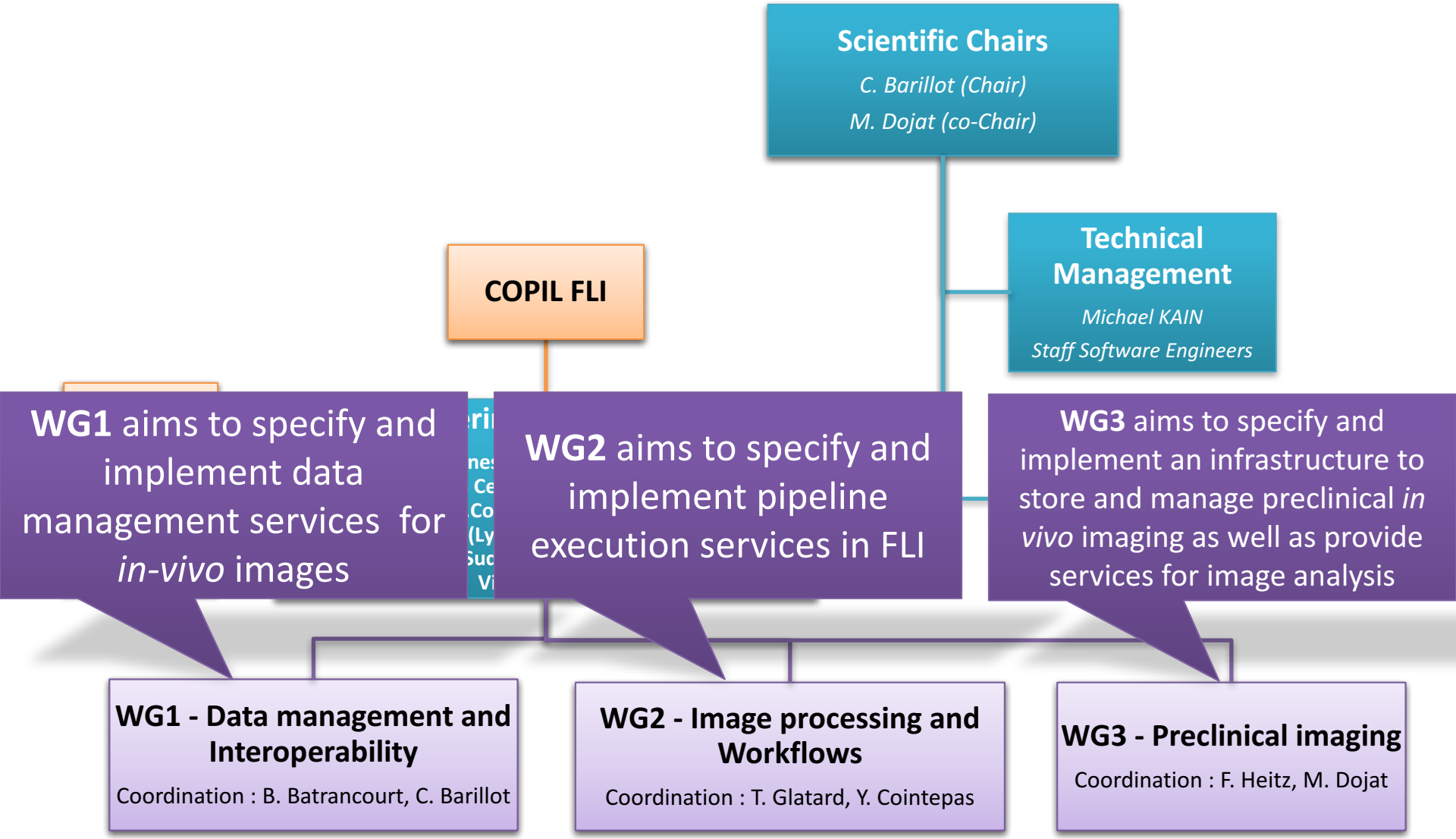
FLI-IAM Node Organization



FLI-IAM Node Organization

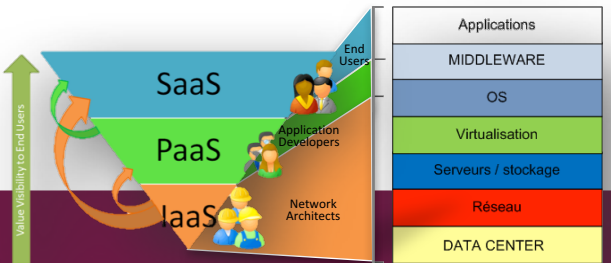


FLI-IAM Node Organization

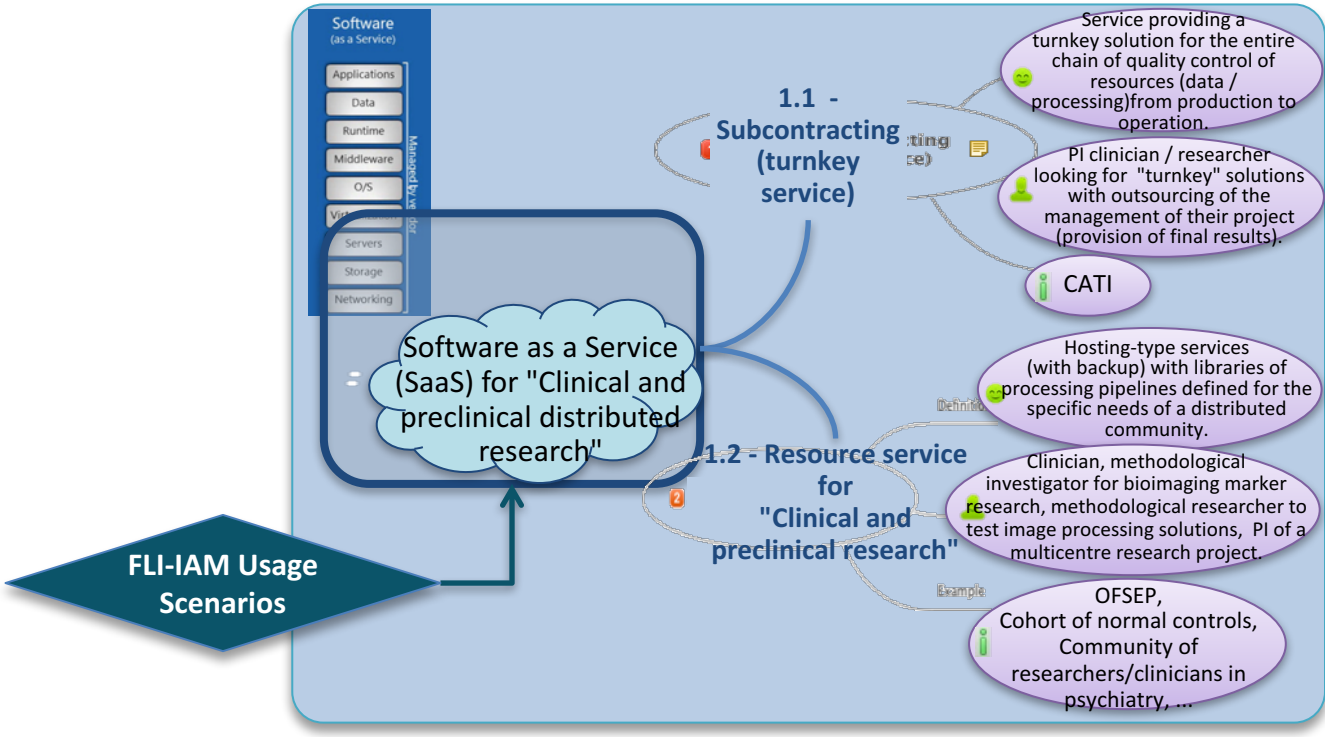


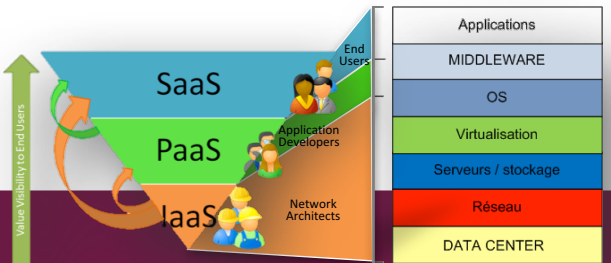
FLI-IAM - Scenarios

FLI-IAM Usage Scenarios



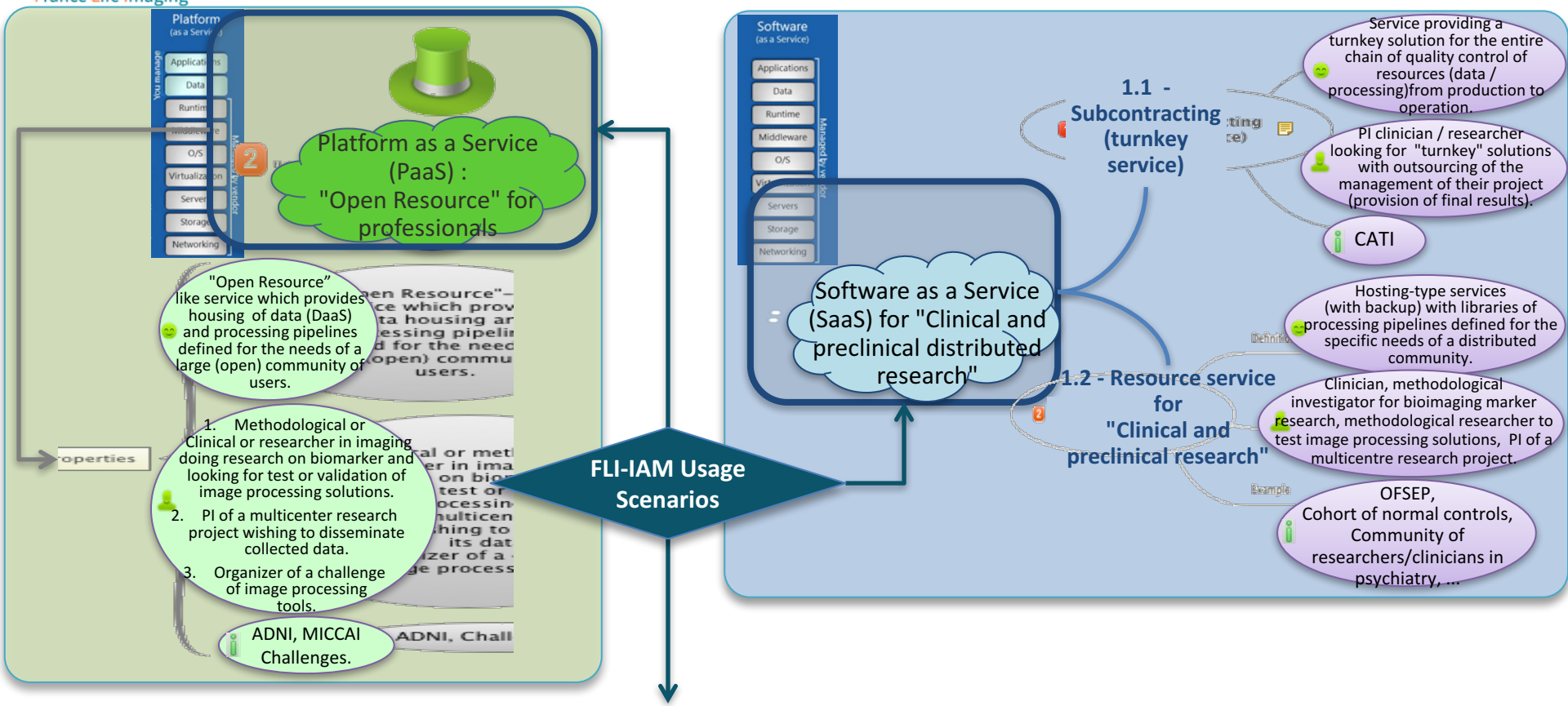
FLI-IAM - Scenarios

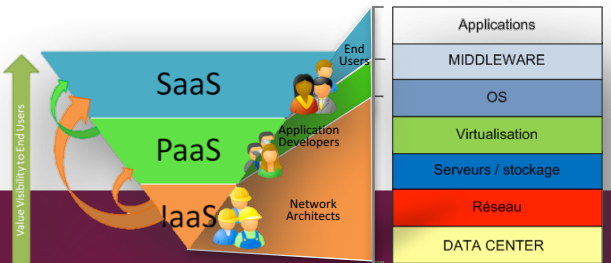




FLI-IAM - Scenarios

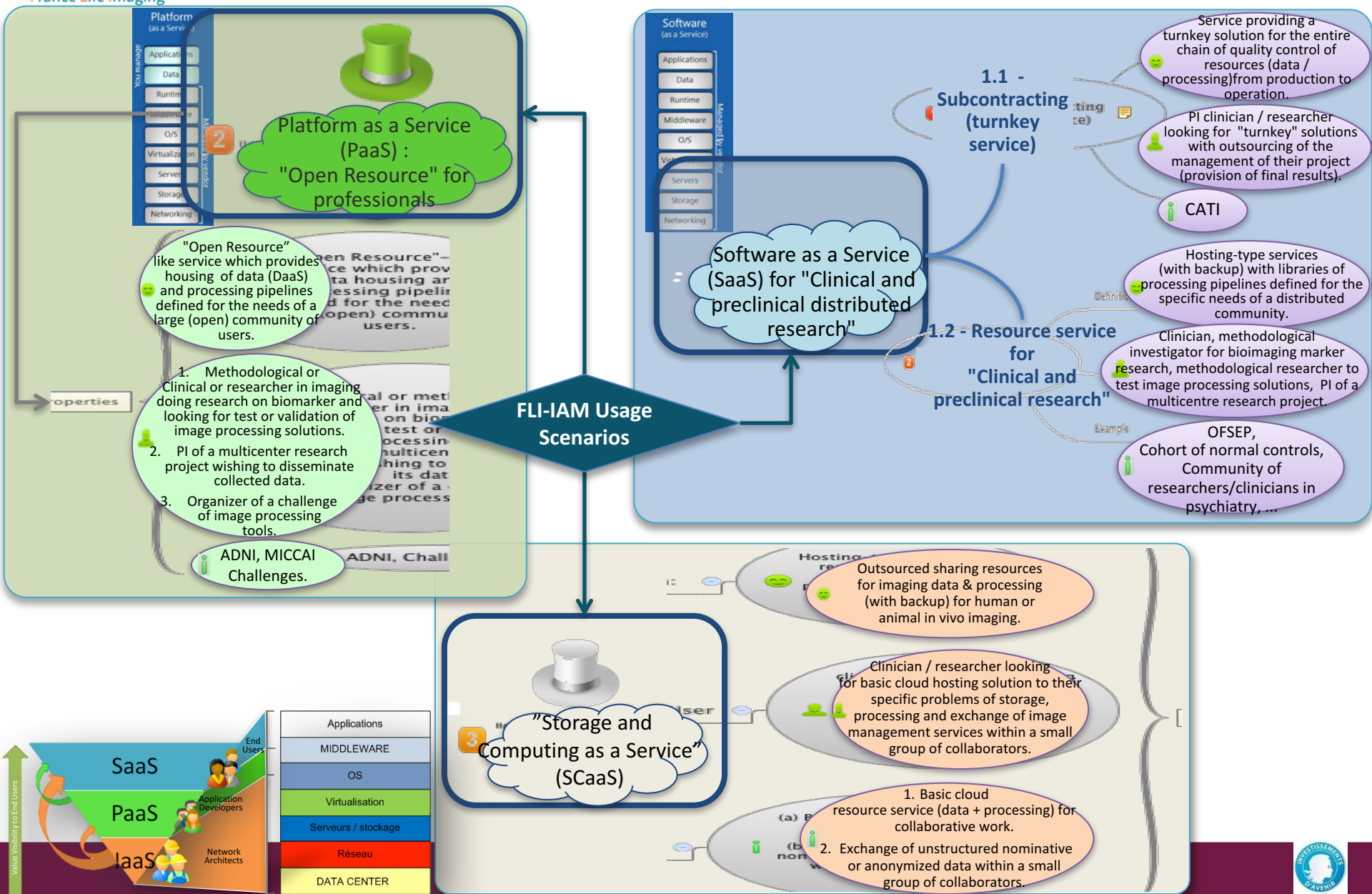
France Life Imaging



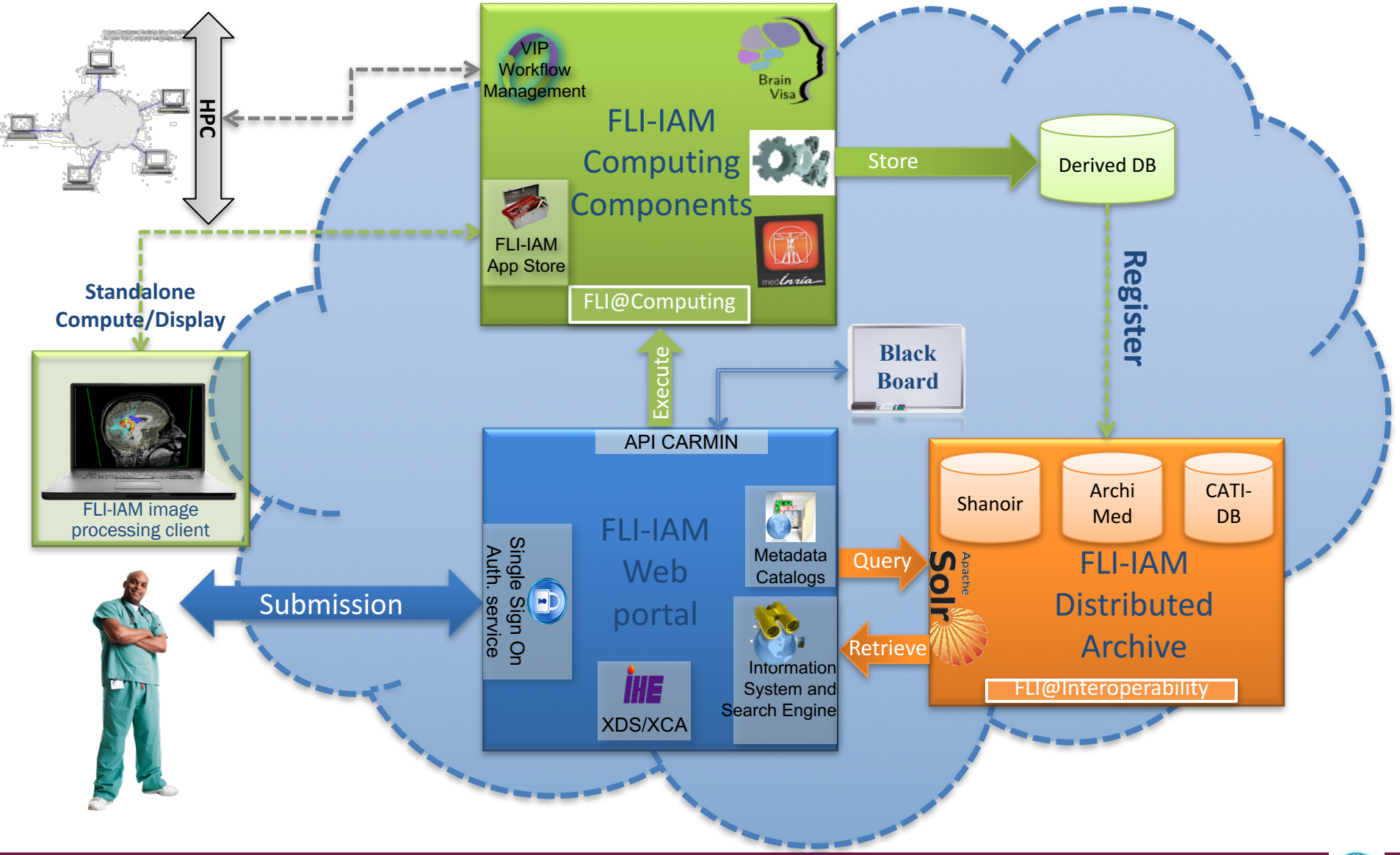


FLI-IAM - Scenarios

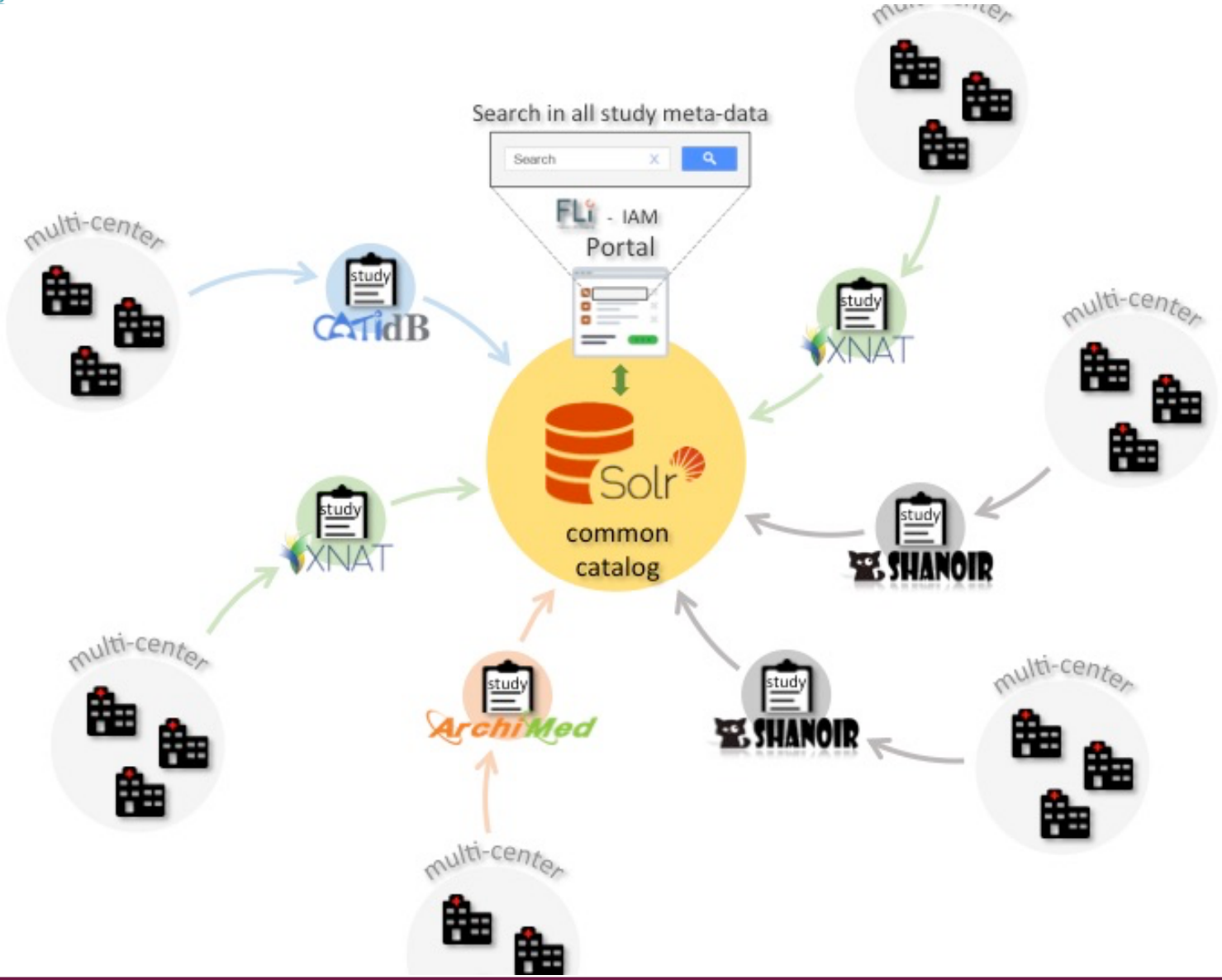
France Life Imaging



FLI-IAM Computational architecture



FLI-IAM – Common Catalog v3



FLI-IAM – CARMIN-API, v0.3

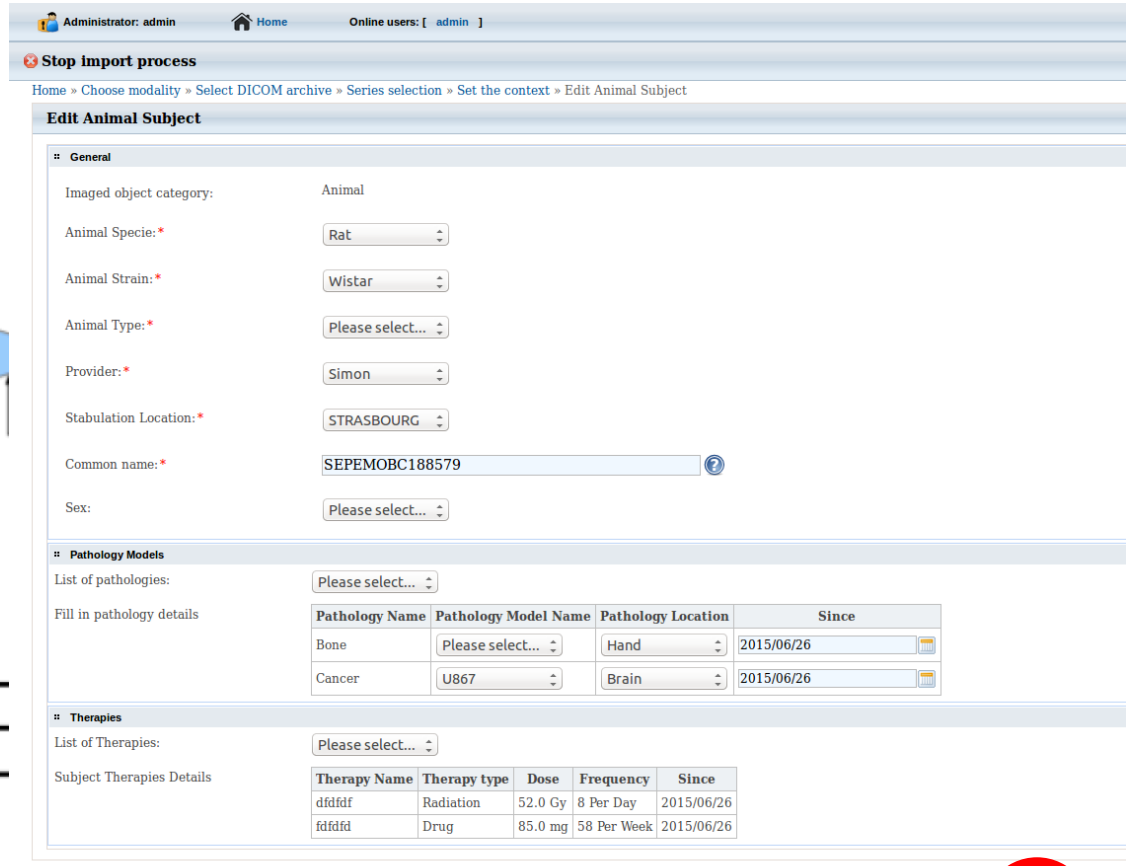
- **CARMIN:**
  - **C**ommon API for **R**eaching **M**edical **I**maging **N**etwork or
  - **C**ommon web **A**PI for **R**e**M**ote pipeline **In**vocation
- **Two web service technologies**
  - SOAP
  - HTTPS/JSON abusively called “REST”
- **https://github.com/fli-iam/CARMIN/**
- **Core API features:**
  - Authentication
  - Platform information
  - Pipelines information
  - Pipeline execution
  - Study isolation (optional)

FLI-IAM Node: Preclinical Imaging

- **68%** of FLI plateforms
- **Bruker scanners**
- **Shanoir-SA**: Extension of Small Animal Imaging to Shanoir system

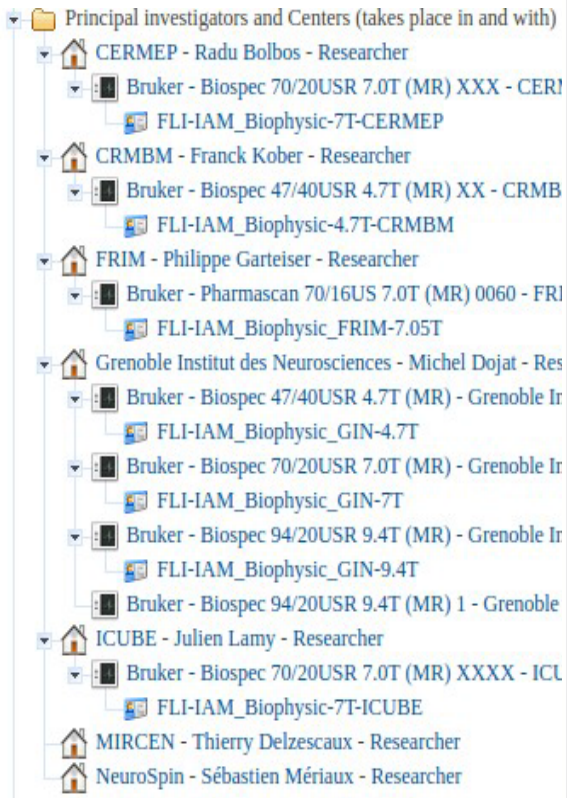
FLI-IAM Node: Preclinical Imaging

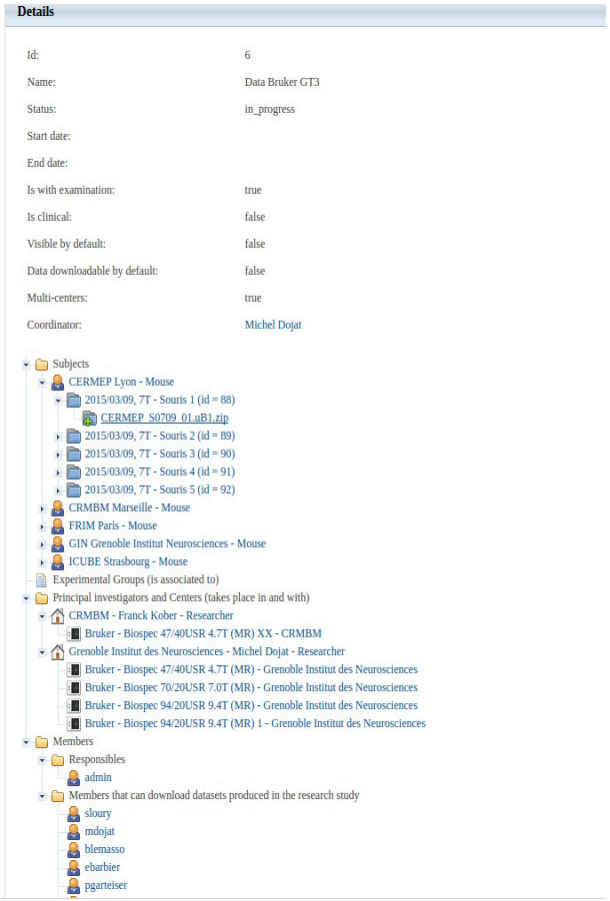
Extension OntoNeurolog ontology

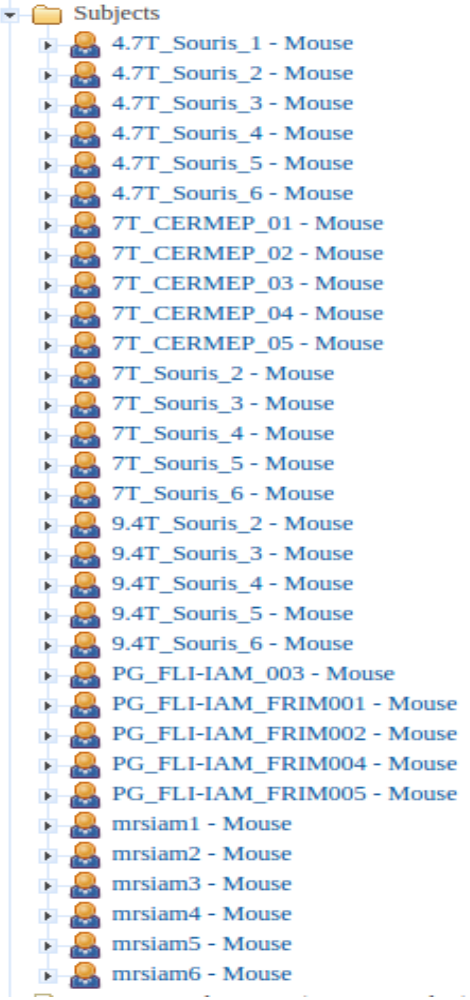


France Life Imaging

SHANOIR







FLI-IAM - Applications

- **3 usage scenarios identified**
  - Software as a Service (**SaaS**)
    - 1.a: turn-key service
    - 1.b: resource service
  - Platform as a Service (**PaaS**)
    - **Show case: organization of two scientific challenges**
  - Storage + Computing as a Service (**SCaaS**)

FLI-IAM - Applications

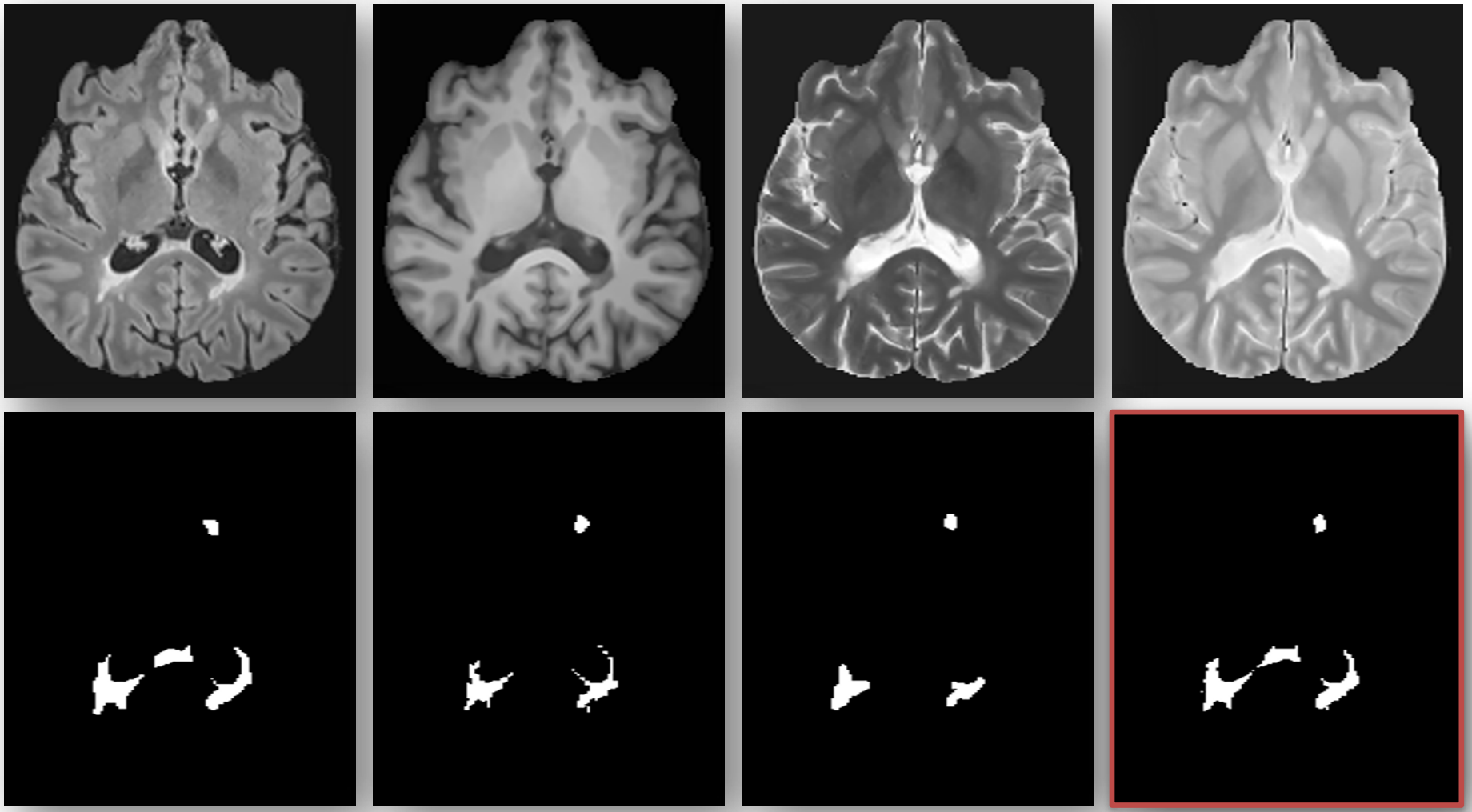
- **3 usage scenarios identified**
  - Software as a Service (**SaaS**)
    - 1.a: turn-key service
    - 1.b: resource service
  - Platform as a Service (**PaaS**)
    - **Show case: organization of two scientific challenges**
  - Storage + Computing as a Service (**SCaaS**)
- **MICCAI 2016**
  - Medical Image Computing and Computer Assisted Intervention
  - Biggest conference on medical image computing
  - **SATELLITE EVENTS: Challenges MSSEG + PETSEG**
  - Ongoing challenges:
    - Scientifics ask for data access and pipeline integration
- **More on https://portal.fli-iam.irisa.fr/services**

FLI-IAM – Real life example – PaaS
Challenge MICCAI 2016 - MSSEG

- **Data**
  - 53 patients (4 different MR scanners @ 1.5 & 3T)
  - Modalities : 3DFLAIR, T2/DP, 3D T1-w, 3D T1-Gado
  - 7 manual annotations per patient
- **Two groups of data:** 15 **training** et 38 **testing**
- **Raw and pre-processed data available**
  - Registration, denoising, inhomogeneity correction, brain segmentation/masking
  - Up to challenger's preference

Center / #exams	Training set	Testing set
GE Discovery 3T (Bordeaux)	0	8
Philips Ingenia 3T (Lyon)	5	10
Siemens Aera 1.5T (Lyon)	5	10
Siemens Verio 3T (Rennes)	5	10
Total	**15**	**38**

FLI-IAM – Real life example – PaaS
MSSEG - Data

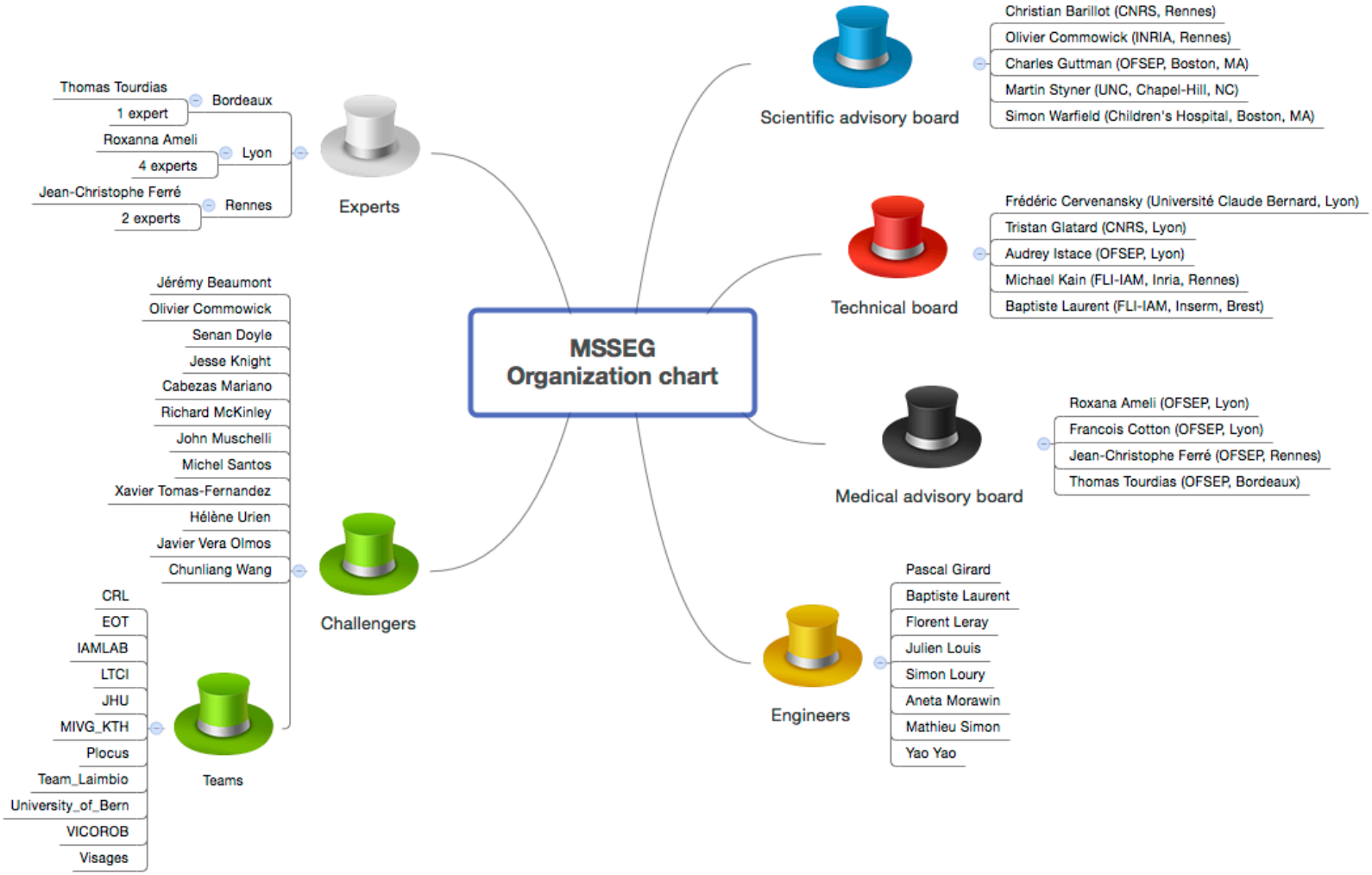


Manual/Automatic Annotations

Consensus
(LOP STAPLE*)

FLI-IAM – Real life example – PaaS

MSSEG - People network



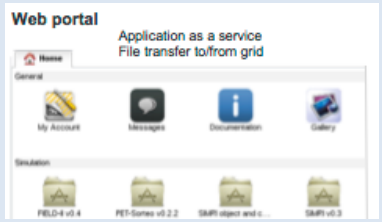
FLI-IAM – Real life example – PaaS Architecture

FLI-IAM Portal

LIFERAY Enterprise. Open Source. For Life.

Processing platform

VIP/Boutiques



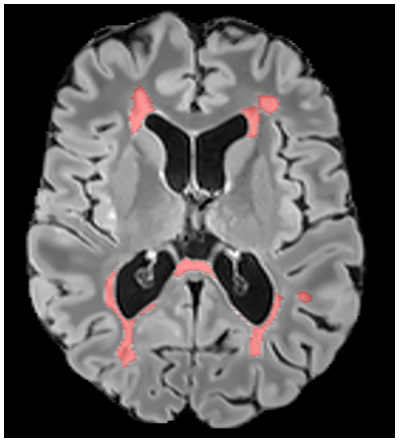
SFTP

DTM

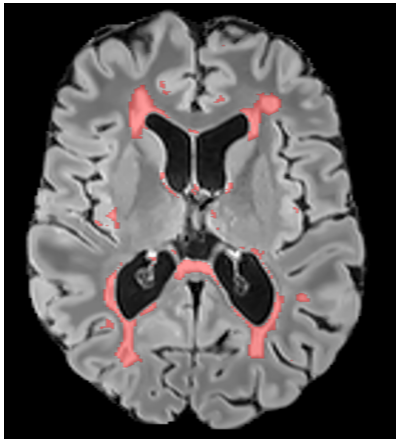
Shanoir

Database for medical images

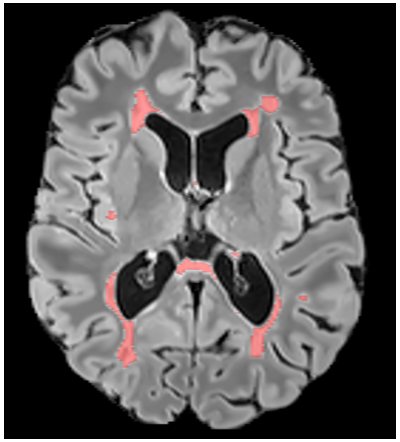
Some Results



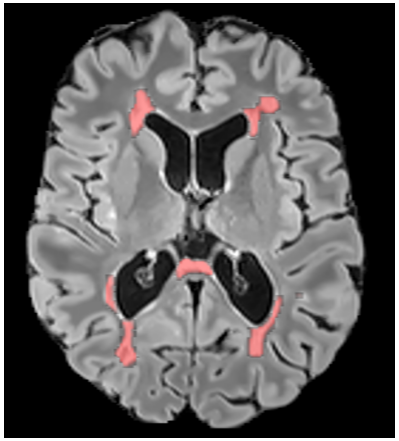
Consensus



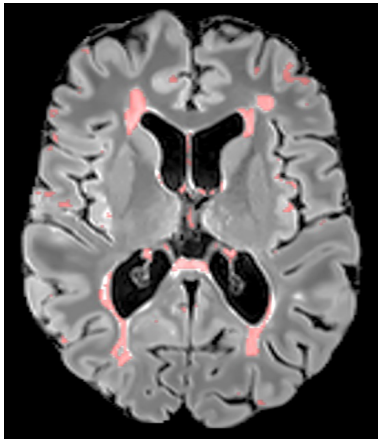
Team 7



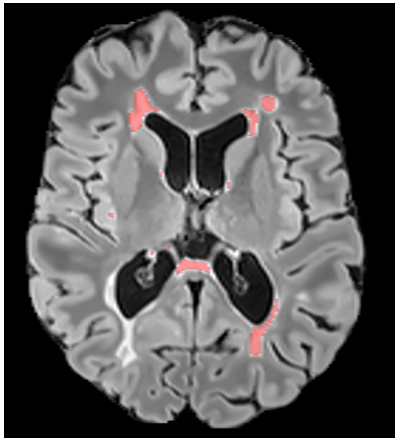
Team 8



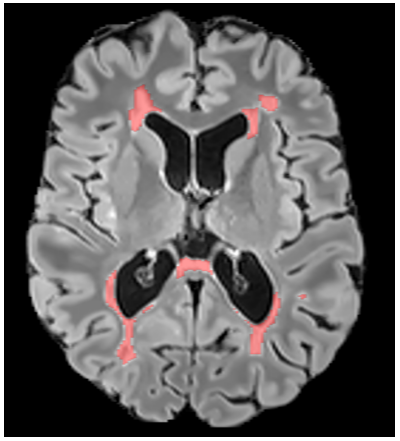
Team 9



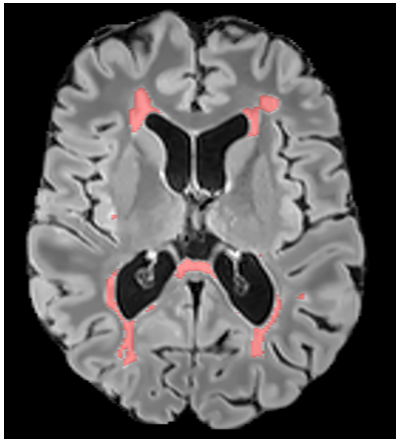
Team 10



Team 11



Team 12

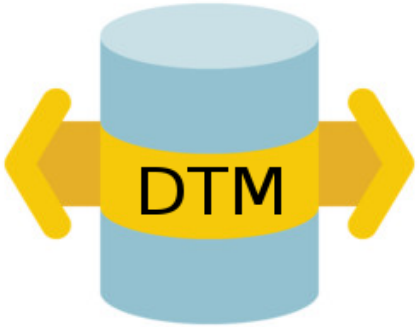


Team 13

Evaluated method	Computation time (on 38 patients)
Team 1	1 day 00h14
Team 2	15h37
Team 3	2 days 00h54
Team 4	04h24
Team 5	15h28
Team 6	2 days 05h40
Team 7	10 days 23h19
Team 8	11h17
Team 9	05h40
Team 10 (*)	5 days 18h37
Team 11	2 days 11h14
Team 12	4 days 16h43
Team 13	2 days 01h03

DTM – Data Transfer Module

Exchange data between platforms
Shanoir_SA ↔ VIP

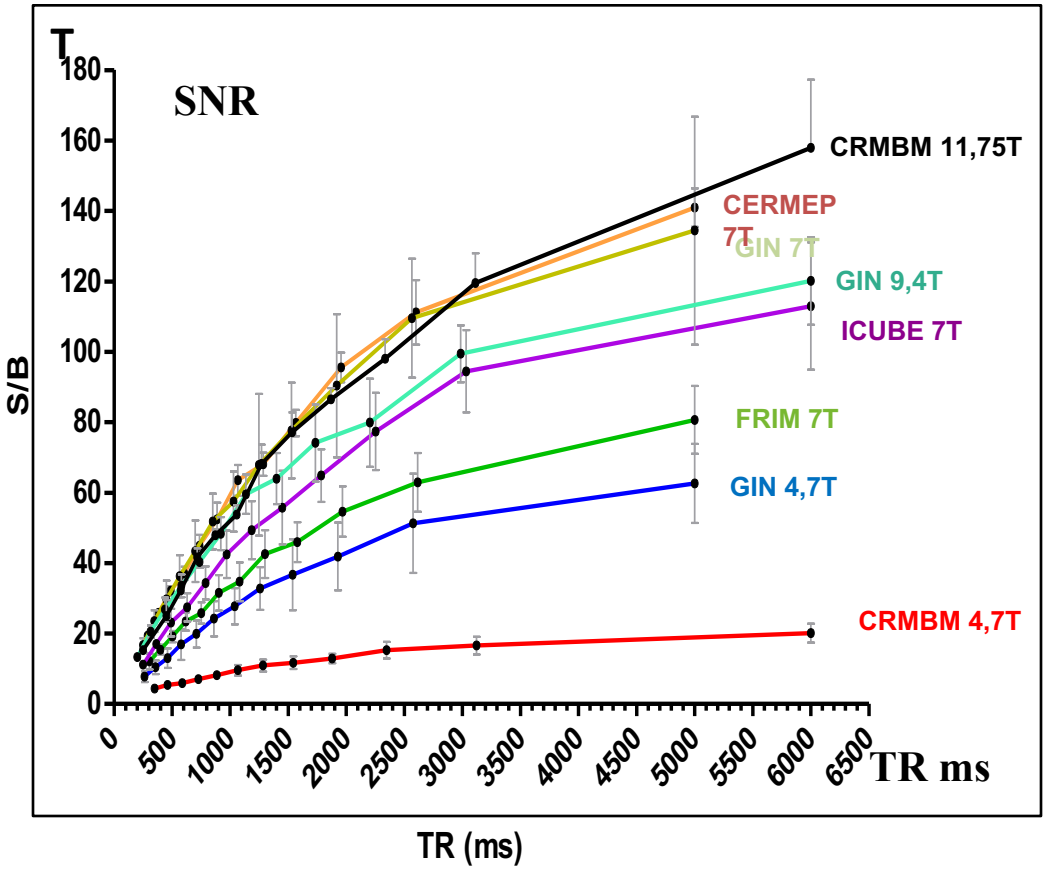


Hosts imaging data/metadata

Exchanges data through webservices

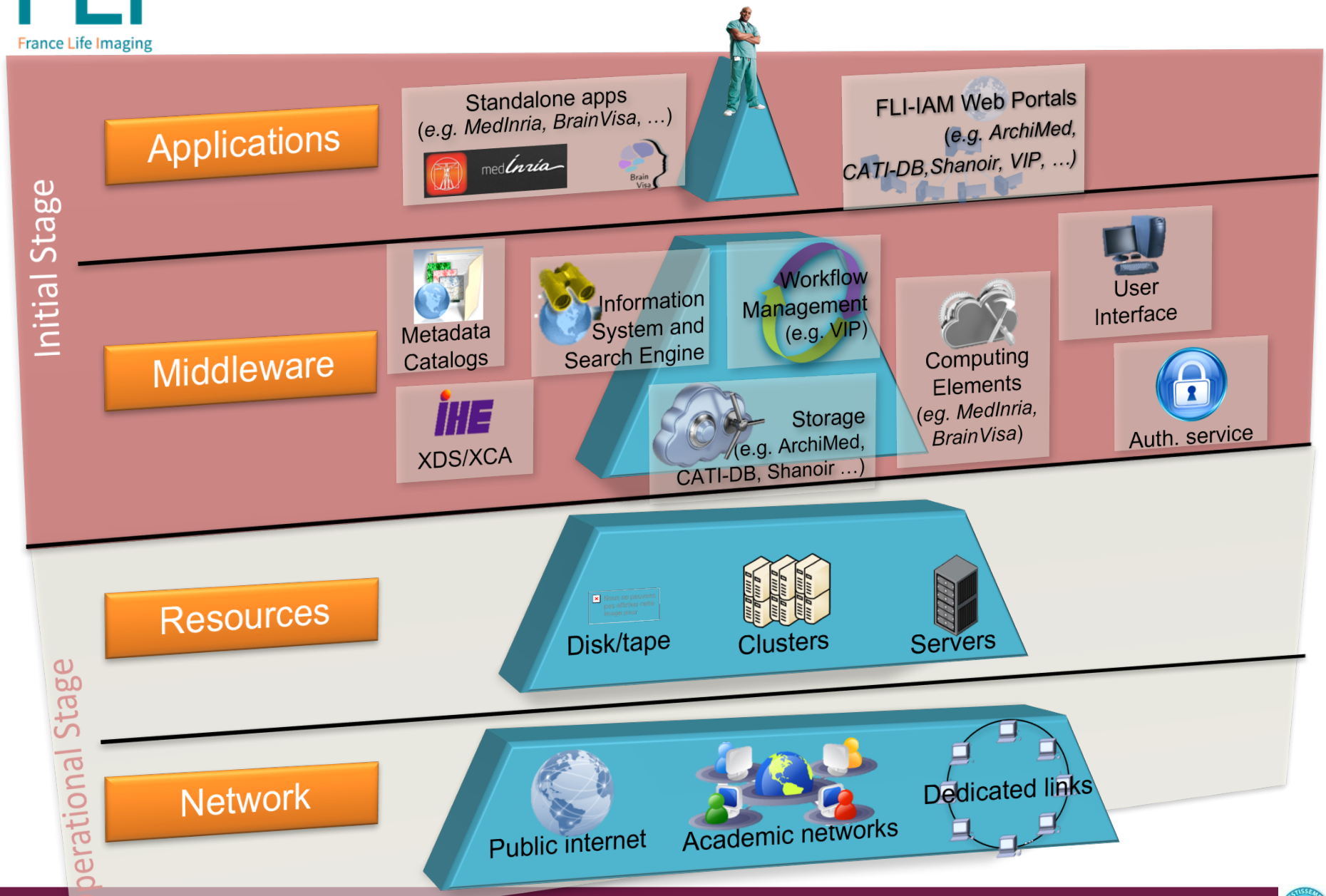
Processes data

- Biophysic UC (5 centers, 5 magnetic field strength)
  - SNR
  - Processing pipelines for T1 and T2 maps computation

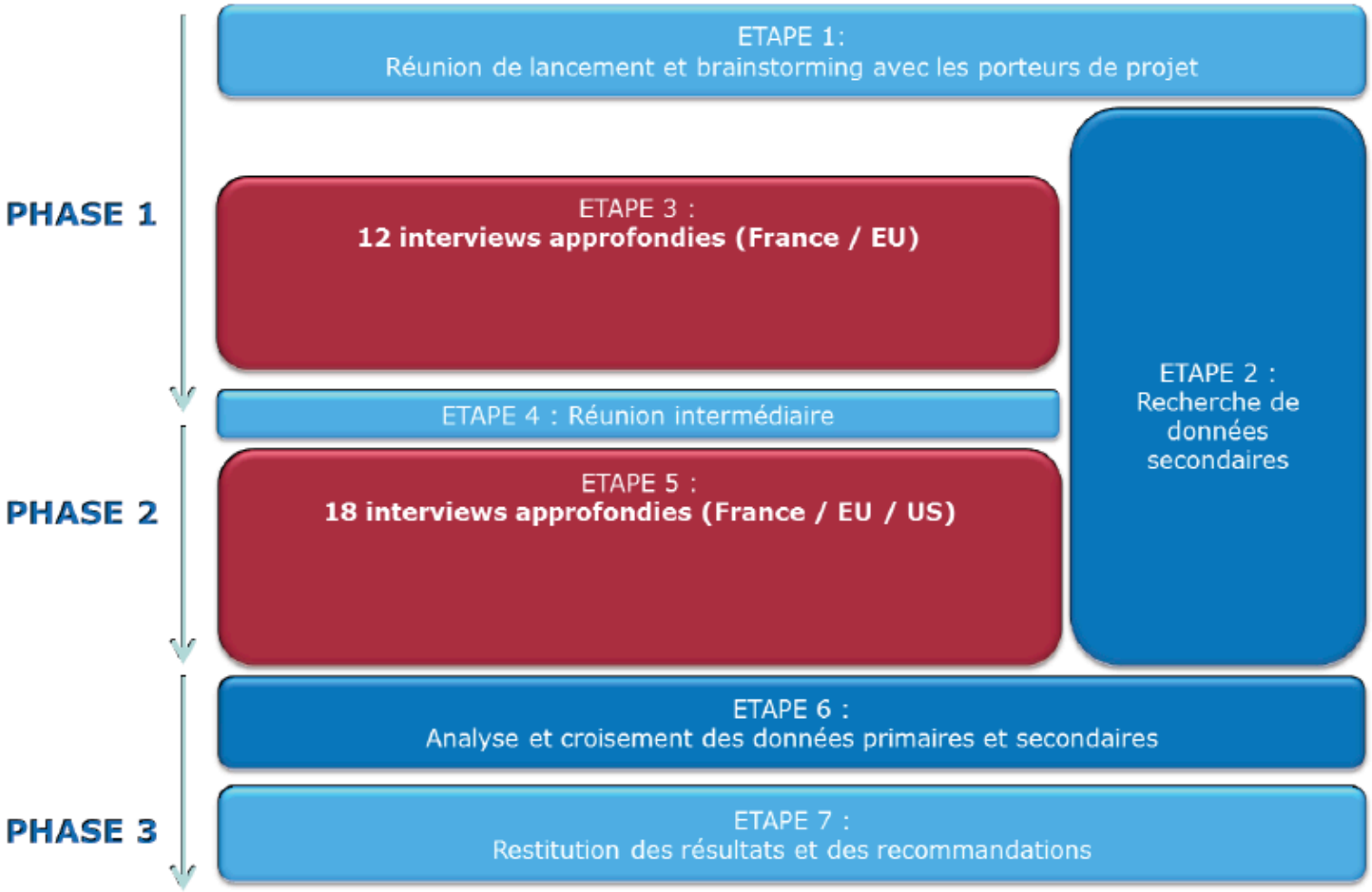


FLI-IAM Stack Architecture

France Life Imaging



FLI-IAM - Market study – project plan

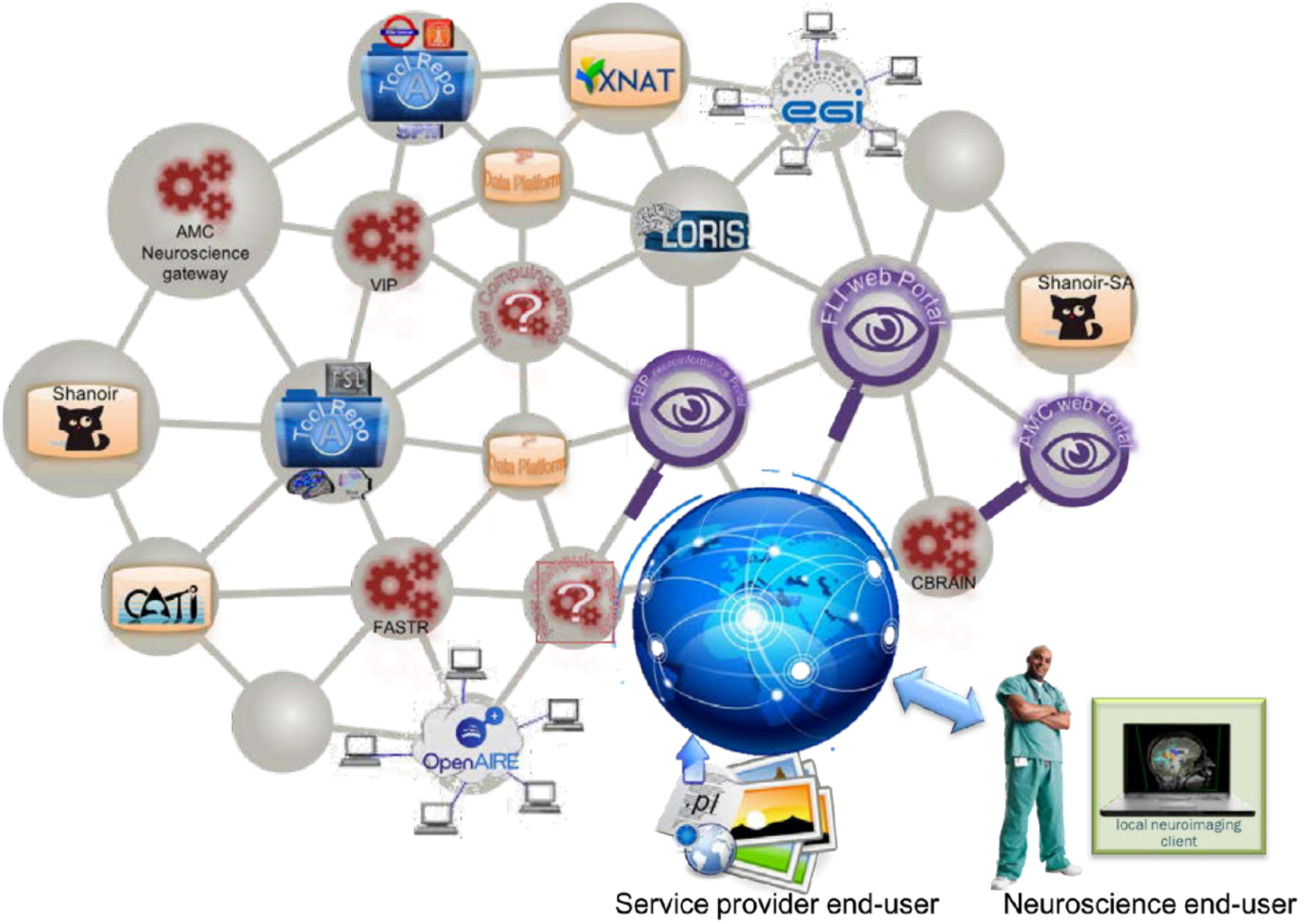


FLI-IAM – Preparation of phase 2

- **Within 2017 we prepare the phase 2**
  - **Market study** in cooperation with The MarkeTech Group (Nantes, California)
  - http://www.themarketechgroup.com/en
  - The first interview phase + intermediary analysis have been finished
  - Timeline:
    - Preparation: target groups + people, interview contents
    - 23.01. – 17.02.: Interviews (12)
    - 17.02.: First step synthesis
    - 20.02. – 17.03.: Interviews (18)
    - **04.04.: Final synthesis**
  - **Target groups for interviews:**
    - **Hospital Informatic managers**
    - **Academic imaging centers**
    - **R&D Pharma**
    - **Radiologists**
    - **CRO**
    - **Medical software editors**

FLI-IAM – Roadmap 2017

- **Higher interoperability**
  - Testing and implementation of CARMIN v0.3 in all involved systems
  - Integrate all systems into Renater FedAuth, unified login
- **End user interface**
  - App Store in portal
  - Simplified and unified data upload
  - Unify access to data and pipelines in portal
- **Common catalog v3, including XNAT**
- **Keep alive challenges MICCAI 2016**
- **Prepare phase 2 and migrate into new structure**
- **Open science solution: partner to H2020 project: OpenAIRE-connect** (https://www.openaire.eu)


Tool Repo
XNAT
egi
Data Platform
AMC
Neuroscience
gateway
VIP
LORIS
Shanoir-SA
FLI web Portal
Shanoir
FSL
Tool Repo
Data Platform
AMC web Portal
CBRAIN
CATI
FASTR
OpenAIRE
.pl
local neuroimaging
client
Service provider end-user
Neuroscience end-user

Risks

- **Local Solutions Interferences**
  - Each local data center provider develops its own solution
  - No compatibility between IAM solutions
  - No Sharing

- **International Competition**
  - COINS (US), LONI (US), CBRAIN (Ca), XNAT (US), HBP-MI (Eu)
  - …….
  - => WS Miccai 2016 => RT Front in Neurosciences 2017

Risks

- **Heterogeneous local solutions**
  - Each local data center provider develops its own solution
  - No compatibility between IAM solutions
  - No Sharing
- **International Competition**
  - COINS, LONI (US), CBRAIN (Ca), XNAT (US), HBP-MI (Eu)
- **Adoption of turnkey solutions from GAFAM, BATX**
  - Azure …..
  - No control from academics

France Life Imaging

Researchers, journals and funding agencies must work together to identify and implement technical solutions that allow the most effective data sharing without greatly increasing the burden on researchers.

Poldrack and Gorgolewski Nat Neuro 2014

The slow adoption of tools and services such as data repositories are indications that **technology alone cannot change scientists' practices; social and cultural factors must also encourage data sharing…. Much sharing is private rather than public**.

Wallis et al. PLOS One 2013

France Life Imaging

Thank you for your attention