

# Sharing Data and Image Processing Pipelines: The Information Analysis & Management initiative

**Michel Dojat**



Gin

Grenoble  
Institut Neurosciences



A photograph of a microscope with a tray of slides in the foreground. The slide in focus shows a grid of red-stained cells. The microscope's objective lenses and eyepieces are visible at the top. An orange semi-transparent box is overlaid on the left side of the image.

# GIN

a center dedicated to  
fundamental and clinical research  
in the field of neurosciences



## OUR MISSION

Understanding the brain  
and developing **innovative therapies**  
for nervous system diseases



## OUR STRATEGY

A **research continuum** from molecules,  
cells to small animals and humans for the  
study of neurological, neuromuscular and  
psychiatric pathologies



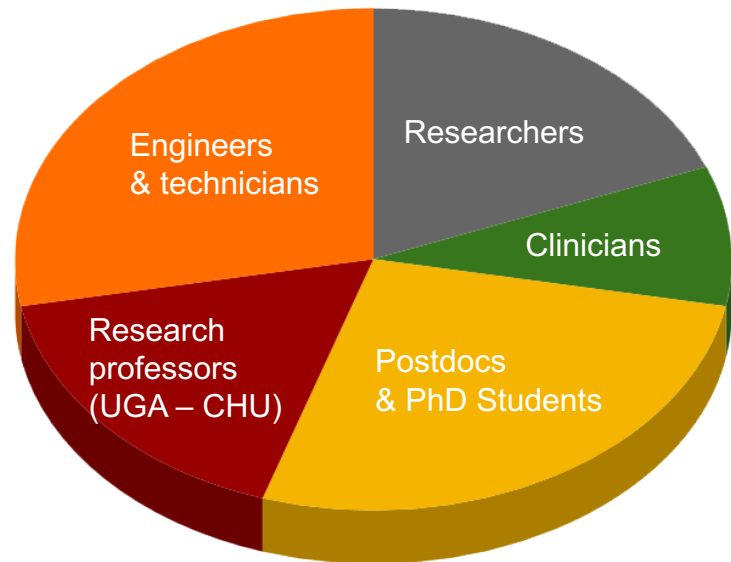
## OUR STRENGTHS

- / Molecular and cellular biology of the neuronal cytoskeleton
- / Predictive models of cerebral pathologies
- / Mapping of neural networks and their alteration in neuropsychiatric disorders
- / Functional neurosurgery and deep brain stimulation
- / Functional MRI coupled with electrophysiology

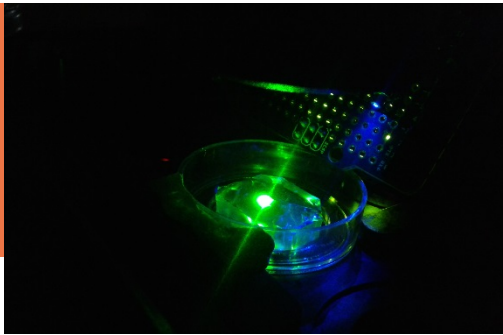


## UP TO 250 NEUROSCIENTISTS

- / 39 Researchers
- / 19 Clinicians
- / 55 Postdocs and PhD Students
- / 34 Research Professors
- / 57 engineers and technicians



# RESEARCH TOPICS AND EXPERIMENTAL APPROACHES



/ **Fundamental neurosciences**

*Cytoskeleton, Intracellular traffic, Synaptic plasticity, Mechanisms studied in normal and pathological conditions (neurobiological diseases, neurodegenerative diseases, myopathies)*



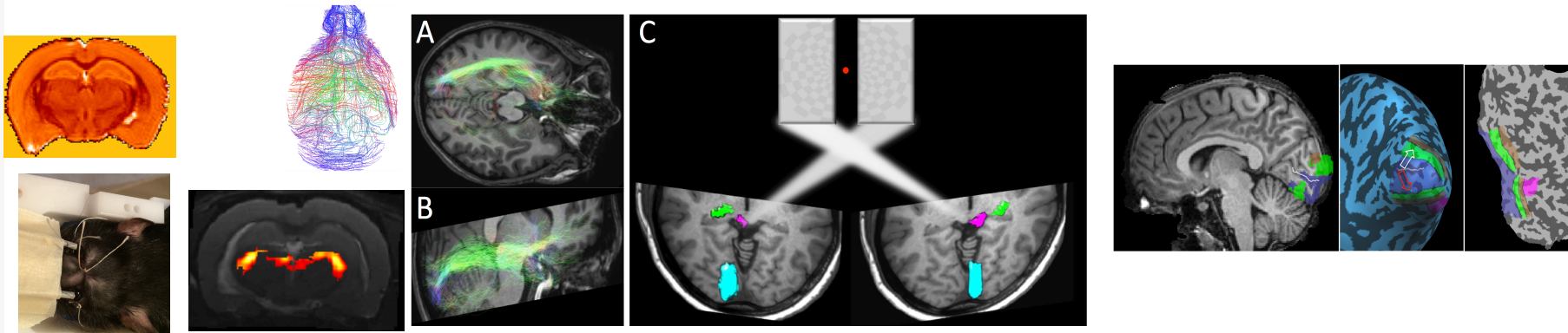
/ **Pre-clinical and clinical research**

*Developing tools and concepts  
Close links with networks such as GREEN, Neuropsynov, NeuroCoG...*



/ **Innovative technologies and treatments**

*Multidisciplinary approaches including human social sciences, and methodological developments (optogenetics, reconstruction of neural networks, electrophysiology...)*



# Sharing Data and Image Processing Pipelines: The Information Analysis & Management initiative

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

---

- **Scientific**

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

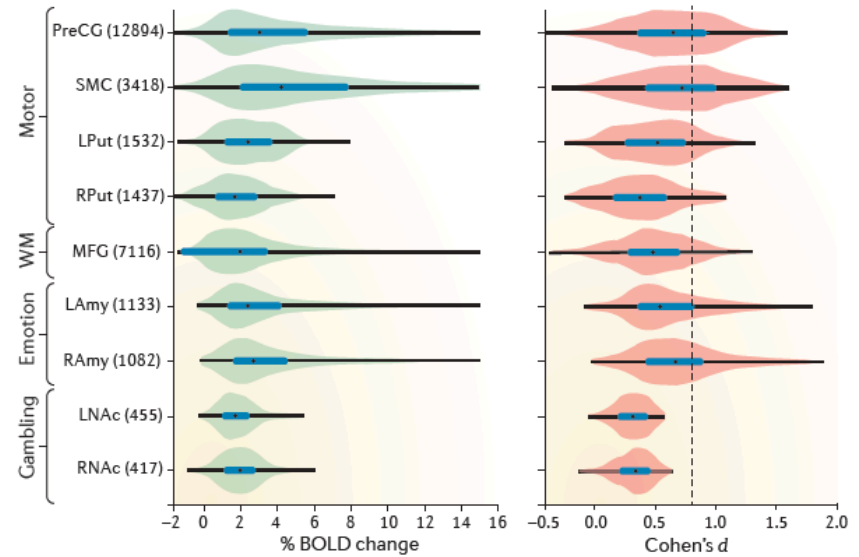
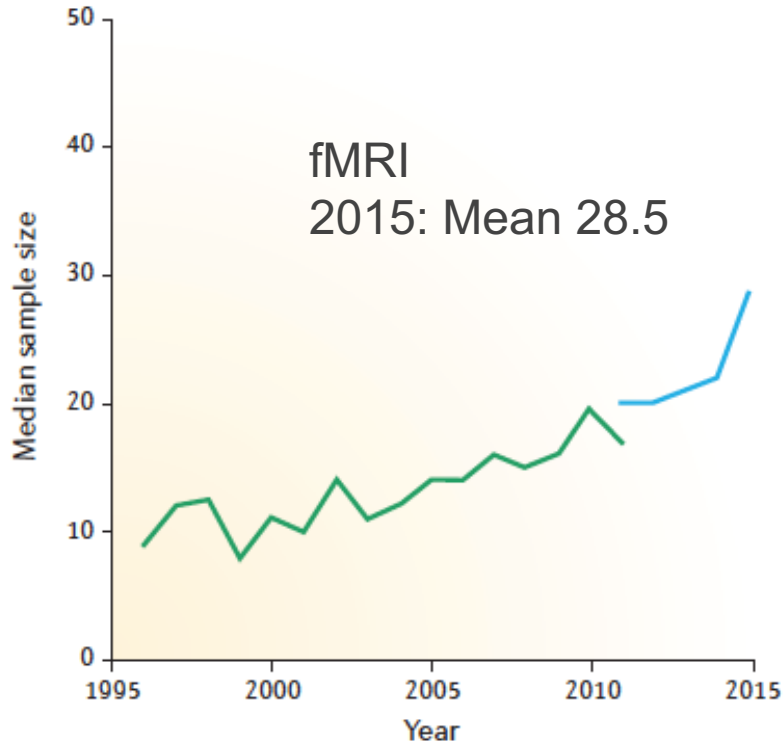
« Power failure: why small sample size undermines the reliability of neuroscience »  
Button et al. Nat. Neurosc 2013

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

Wicherts et al. PLOS One 2011

- **Rarely replicated so false solutions persist**

# Old fashion



[Poldrack Nat Rev Neurosc 2017]

**Large cohort, Re-analysis, Replication, MetaAnalysis**

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

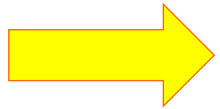
« crowd science » Franzoni & Suerman Res Policy 2014



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

- **Economic**

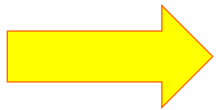
Data Acquisition: e.g. neuroimaging 300€/H 30 subjects x 2 groups 18 K€ ...  
Many efforts for image processing tools development and validation  
don't replicate existing data and tools ....  
Select the best processing pipelines for a target application



**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**

# Paradigm Shift: BIG DATA, OPEN SCIENCE + AI

---

## **Dermatologists vs CNN**

**[Esteva et al Nature 2017]**

127463 biopsies for training

1942 for validation

Inception v3 (GoogleNet, using several replicas on Nvidia Titan X Gpu)

## **Ivy Glioblastoma atlas**

**[Puchalski et al Science 2018]**

41 patients, 42 tumors, 440 tissue blocks, 270

transcriptomes

11500 annotated H&E images

23000 ISH images (**400 Gb/image**), 400 MRI scans

**[Poplin et al Nat Biomed Eng 2018]**

## **Cardiovascular risks from retinal fundus photos**

284335 patients (EuePACS + UKBiobank)

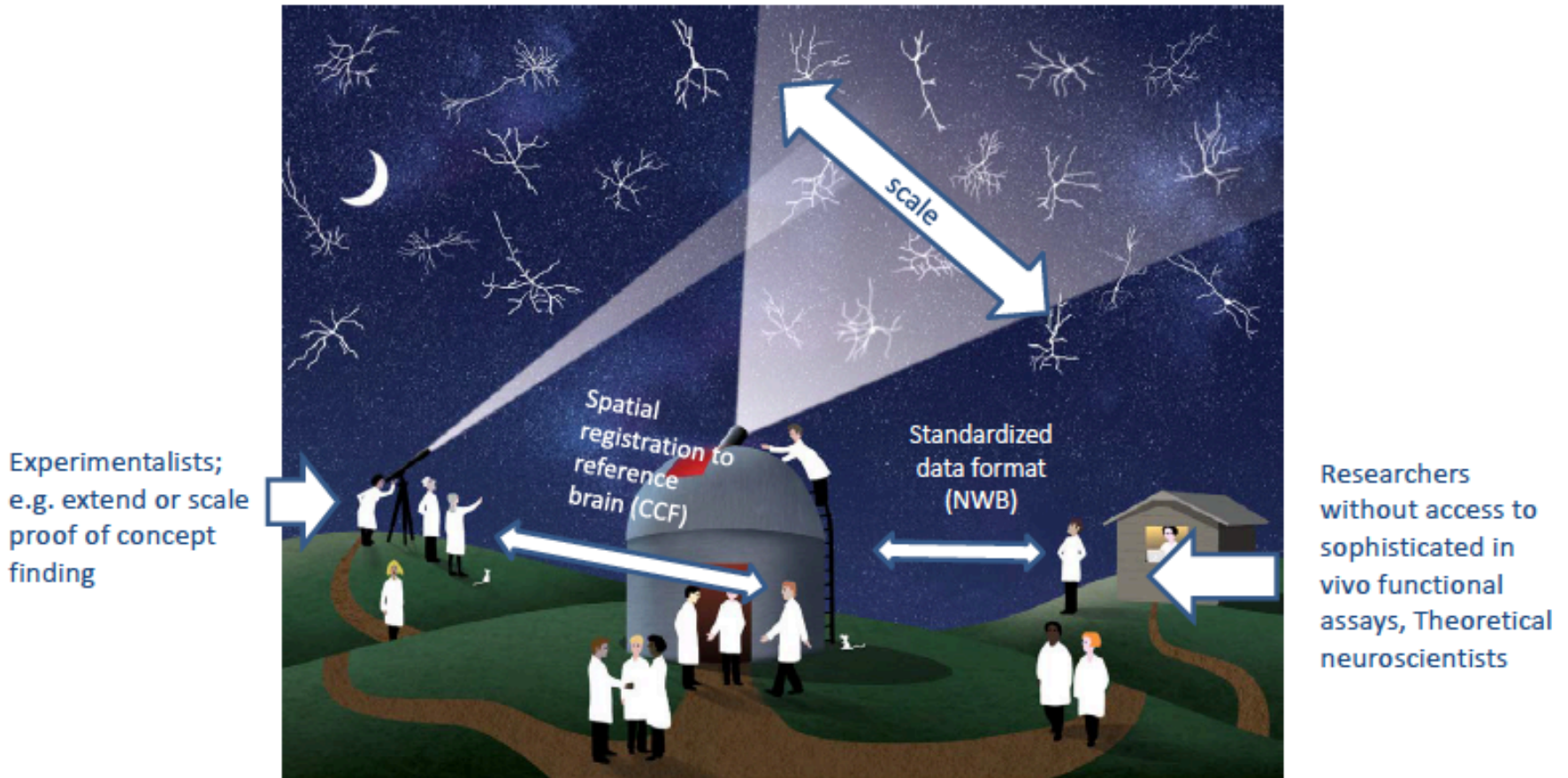
12096 and 999 for validation

Inception v3 (GoogleNet, using several replicas on Nvidia Titan X Gpu)

22 millions of parameters to be estimated


# New (BIG DATA) model of neuroscience

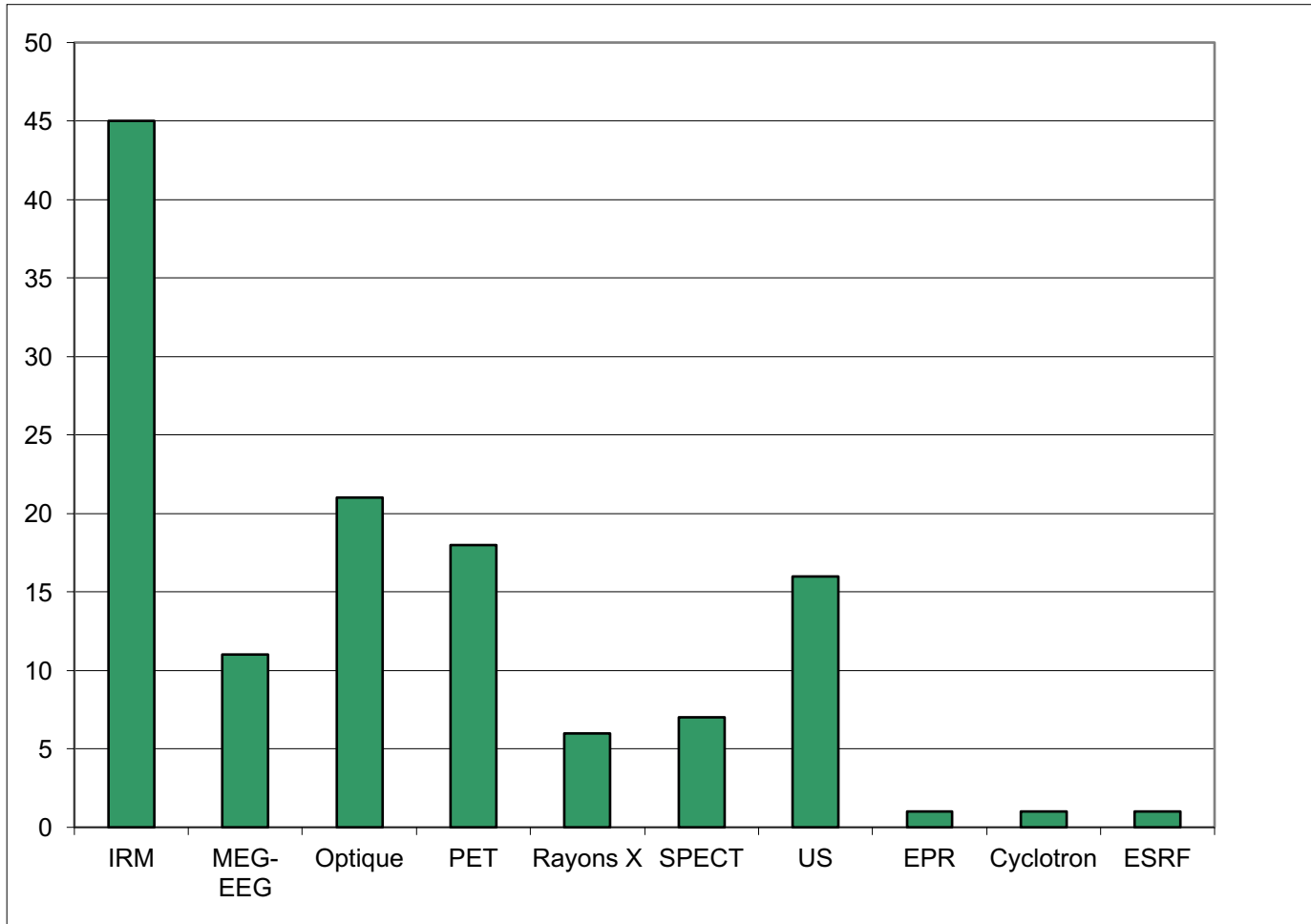
How from the activity of millions of neurons distributed in several brain areas emerge simple percepts and how they are linked to emotion, motivation or action?



courtesy of Mike Hawrylycz

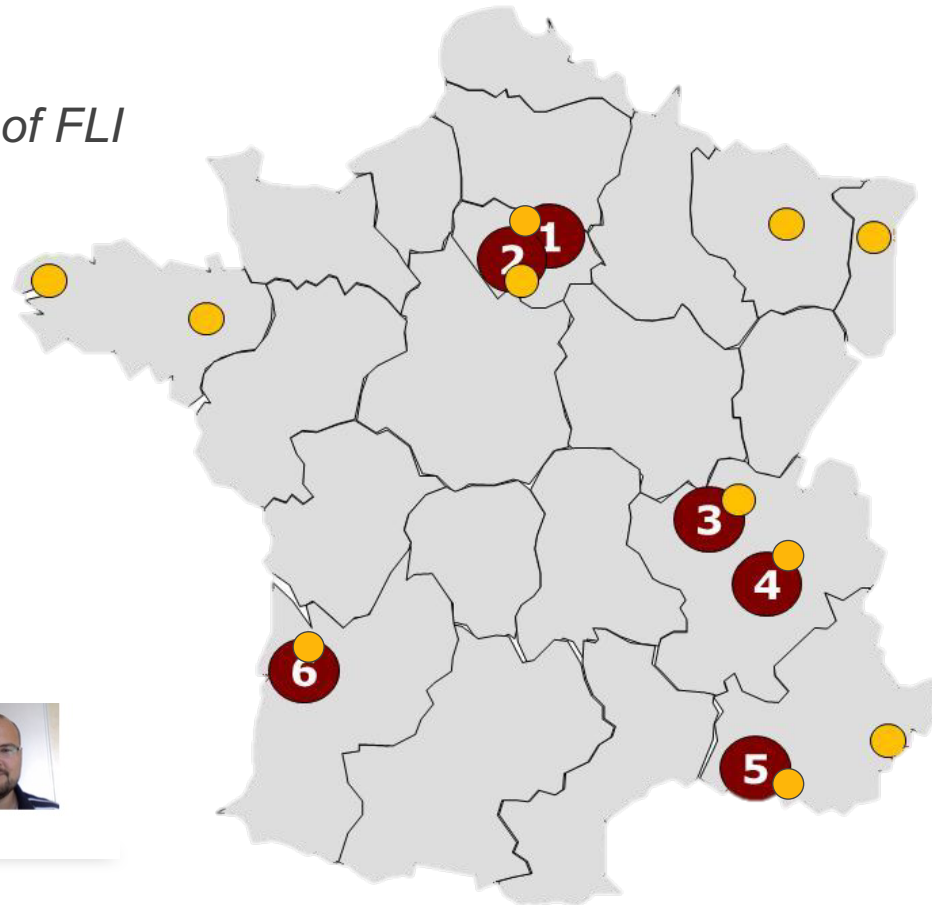
## Infrastructure project for a coordinated and harmonized network of *in vivo* imaging in France.

- to coordinate research activities and combine the skills to push the current technological barriers
- to provide scientists a complete range of imaging technologies (150 imaging systems) **and integrated services.**
- open to collaborations with industrial partners.
- complements France BioImaging for *in-vitro* Imaging (FBI), the French contribution to  EURO-BIOIMAGING on the ESFRI roadmap.
- 37 M€



32 % Clinical  
68 % Preclinical

*The Numeric infrastructure of FLI  
(5 M€)*



● 6 physical nodes

● 1 transversal node for Information Analysis and Management



**Christian BARILLOT**  
*FLI-IAM Chair (Rennes)*

**Michel DOJAT**  
*FLI-IAM co-Chair (Grenoble)*

**Michael KAIN**  
*FLI-IAM Technical Manager (Inria)*

# FLI-IAM: The Numeric Infrastructure

---

## Our Vision

=> Targeted Users

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



# FLI-IAM Market Study

THE MARKETECH GROUP



Final Report

## Project IAM

Prepared for:

Michael Kain  
France Life Imaging (FLI)  
IAM - Technical Manager

Mai 2017



# Interviews

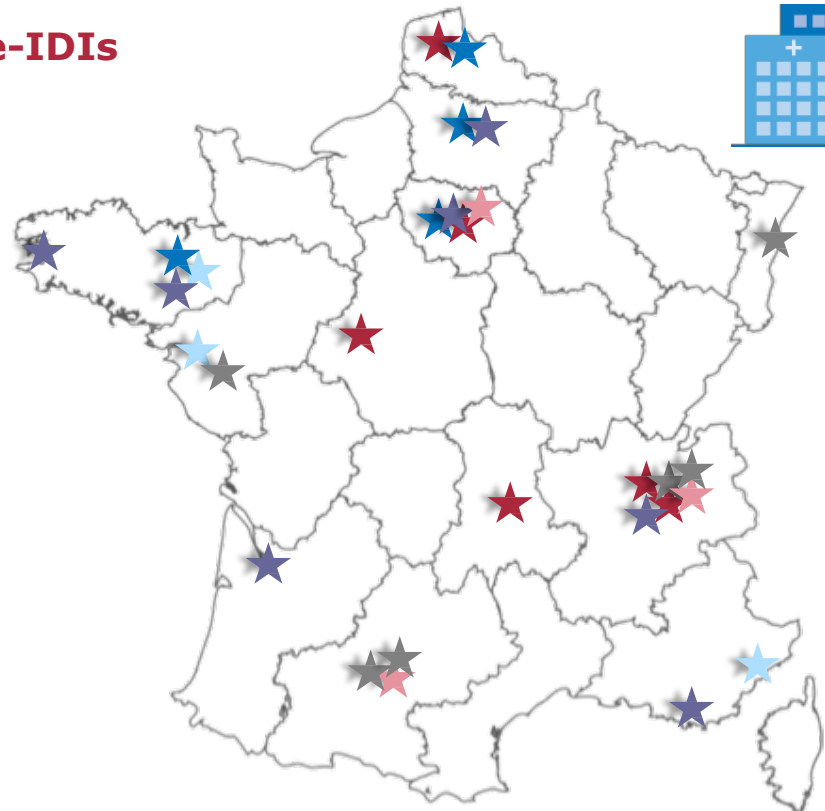
## 30 persons



Tele-IDIs



Segments	Number
★ Clinical imaging	4
★ Preclinical imaging	6
★ Clinical Research Organization	6
★ DSI	4
★ R&D Pharma	3
★ Software solutions	7
<b>TOTAL</b>	<b>30</b>



## Your ideal

- Delocalized and virtual storage and archival solution
- Shared at a **national level**
- Gathering clinical (anonymized) data, imaging data & meta-data

- **Certified by regulatory agencies**  
(e.g. CNIL, ANSM, GDPR, ...)

- Manage various formats (Dicom, Nifti, Bids, ...)

- Guarantee:
  - **Data security**
  - **Stability of the structure**
  - **Health data hosting certification**
  - **Independence from services providers**
  - **Governmental approval of the structure(trustee)**



## At start institutional researchers



- Target: **Academic Users**
- FLI-IAM improves multi-centre studies:
  - **Large medical imaging** repositories
  - **Sharing preclinical data**
  - These repositories allow:
    - Evaluation of image processing software
    - Large statistical studies
  - Added value: a catalogue of **certified software**
- **Preclinical imaging** (small animal), neurology and oncology seem the more interesting applications (large datasets generation)

# FLI-IAM: A two-steps process

---

- **Initial stage (2013-2017)** => 15 engineers
  - Set up of the infrastructure hardware & software : specifications function of various categories of users, roadmap
  - Pilot phase on targeted *in vivo* imaging application for clinical and preclinical research
- **Operational stage (2018 and onward)**
  - Provide an operator
  - Extend the infrastructure to broaden the range of clinical and preclinical research application domains
  - Extend the range of software integration by providing open interoperability standards with conformal statements

# FLI-IAM: The Numeric Infrastructure

## Strengths: Based on high technological expertise and experience

- From data management solutions



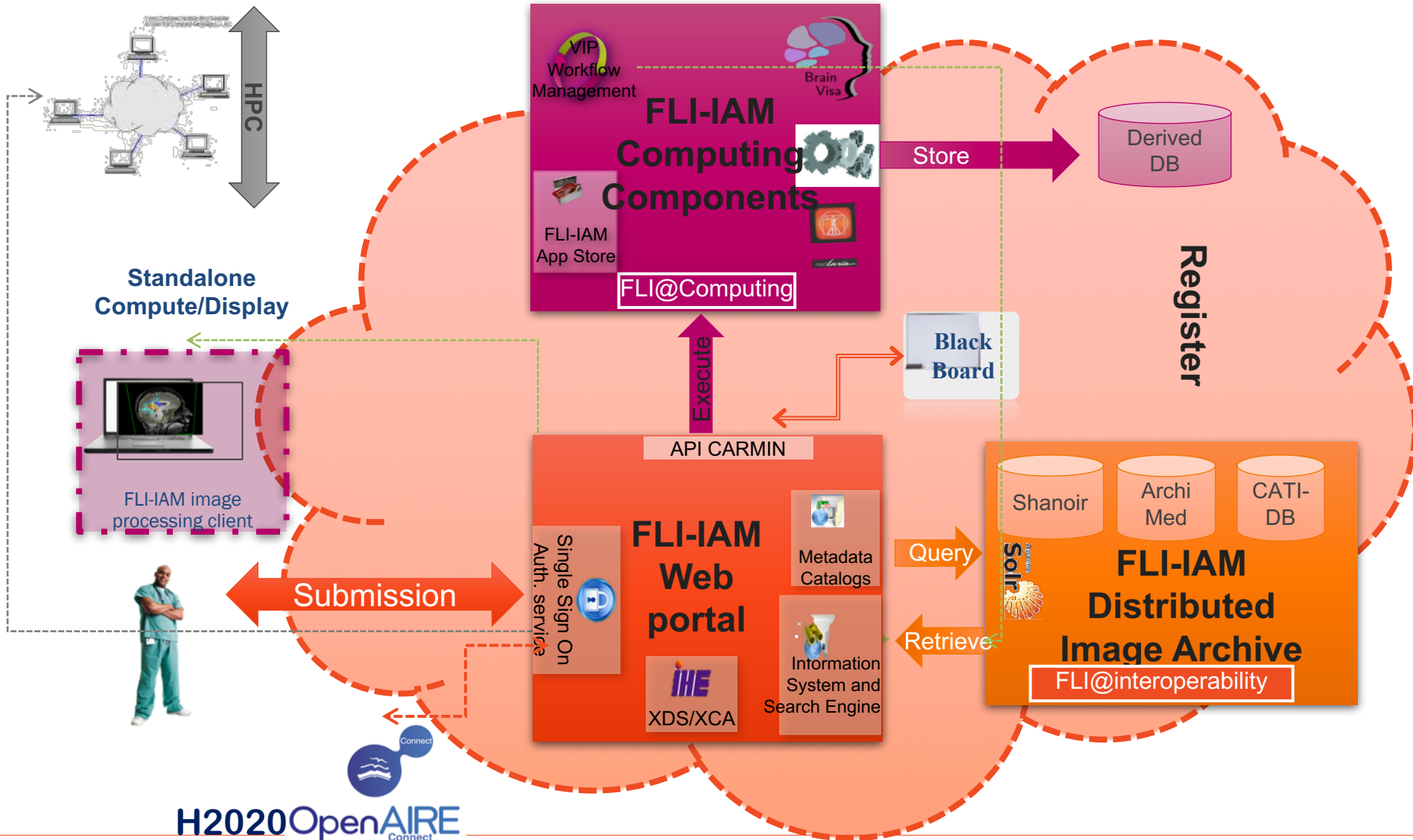
- From medical image processing solutions



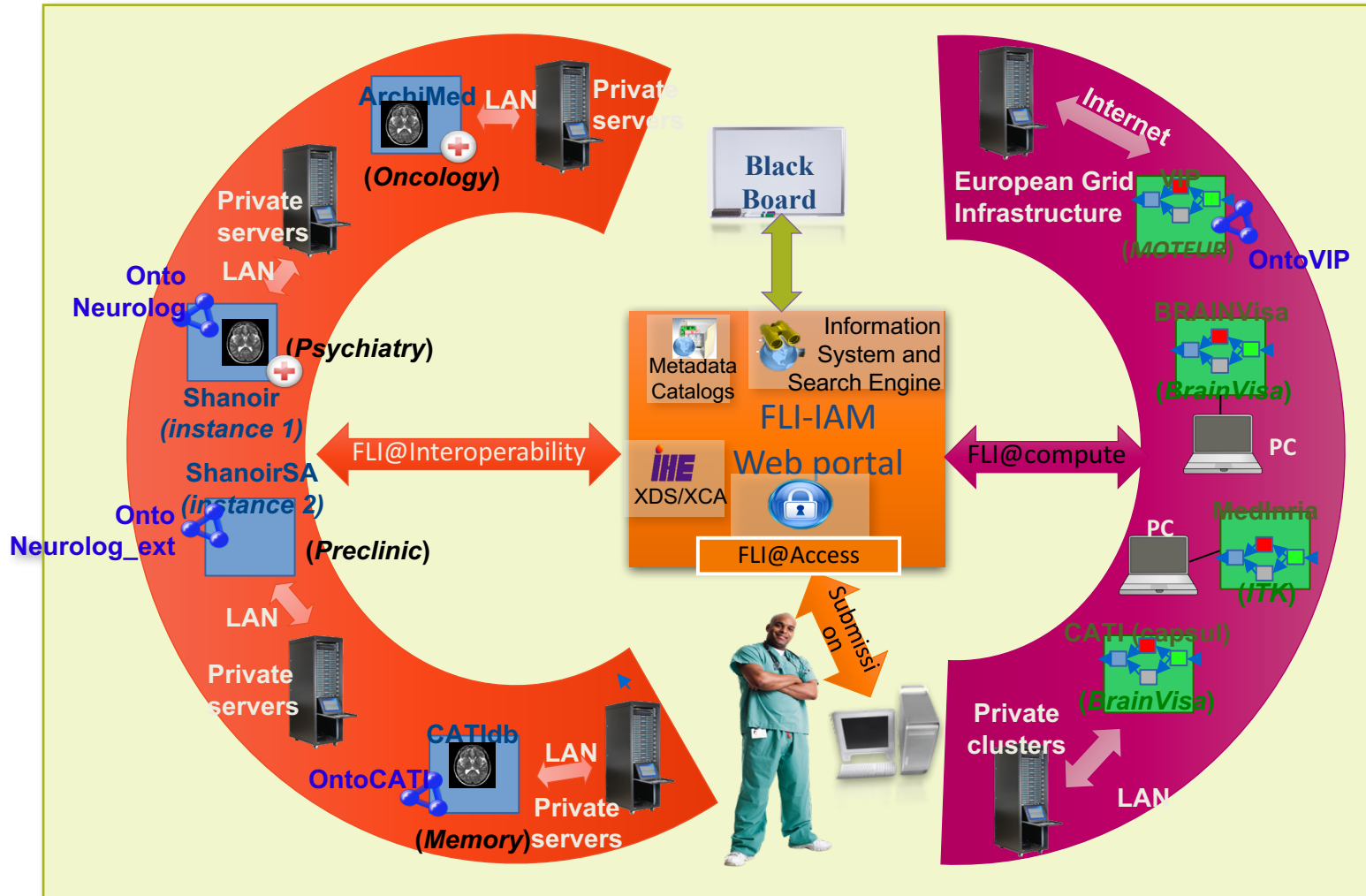
- From large national clinical cohorts: CATI, OFSEP, ...



# FLI-IAM: The Numeric Infrastructure



# FLI-IAM: The Numeric Infrastructure



# FLI-IAM: Key Milestones

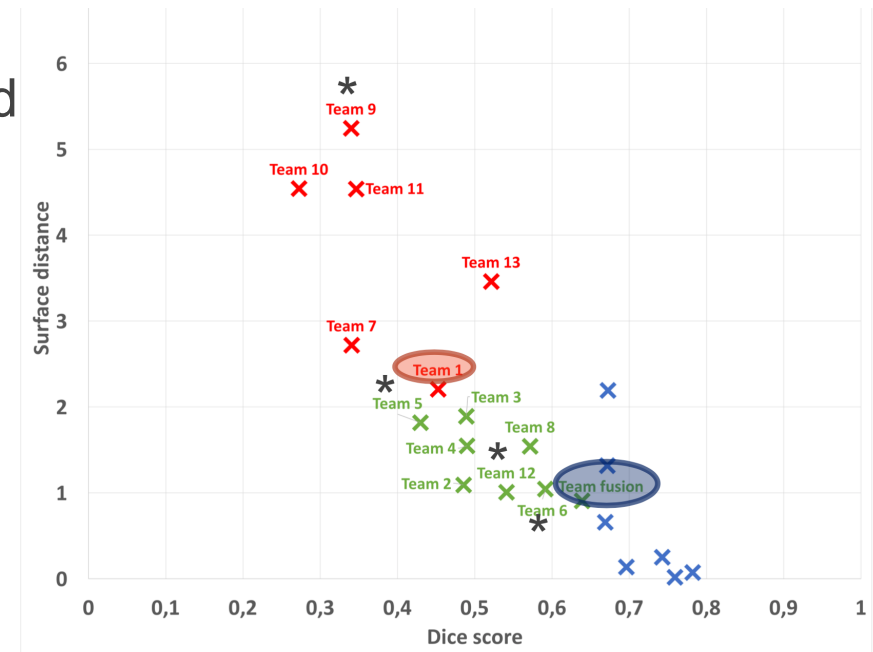
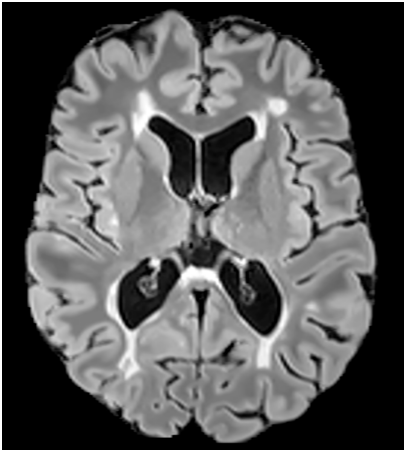
---

- Interoperability with national & international (Xnat) existing solutions
- Link with VIP for pipeline execution
- Link with local environments
- Portal for Data & Tool execution
- H2020 OpenAIRE-connect for Open science
  
- Two MICCAI-challenges: MRI for MS and PET for oncology
- Small animal platform



# FLI-IAM: Challenge MICCAI 2016 MS Segmentation

- 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 (>370 annotations overall)**
- Two groups of data: 15 training et 38 testing
- Raw and pre-processed data available
  - Registration, Denoising, Inhomogeneity correction, Brain segmentation/masking
- **13 Challengers worldwide**  
(machine/deep learning + model-based)



Commowick, O. et al. Nature Scientific Reports, 2018

# FLI-IAM: For Preclinical Studies

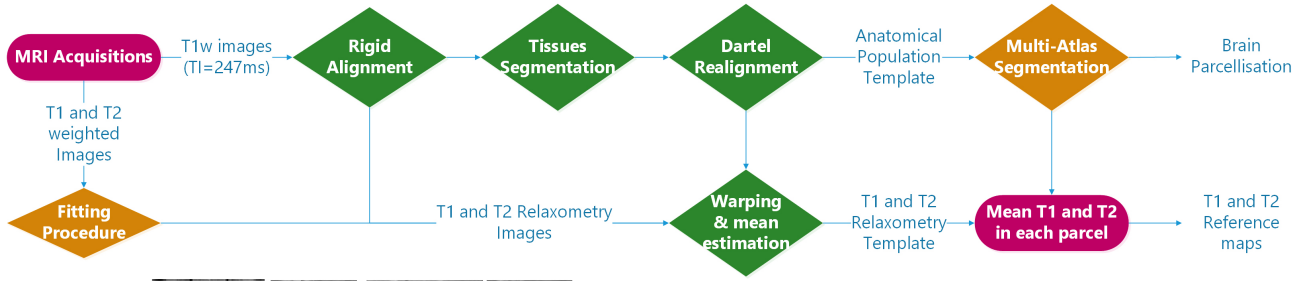
- FLI : 68 % preclinical studies
- Similar needs that for human studies
- Still in infancy
  - => Appning WS (Paris, June 2018)
- Bruker equipment



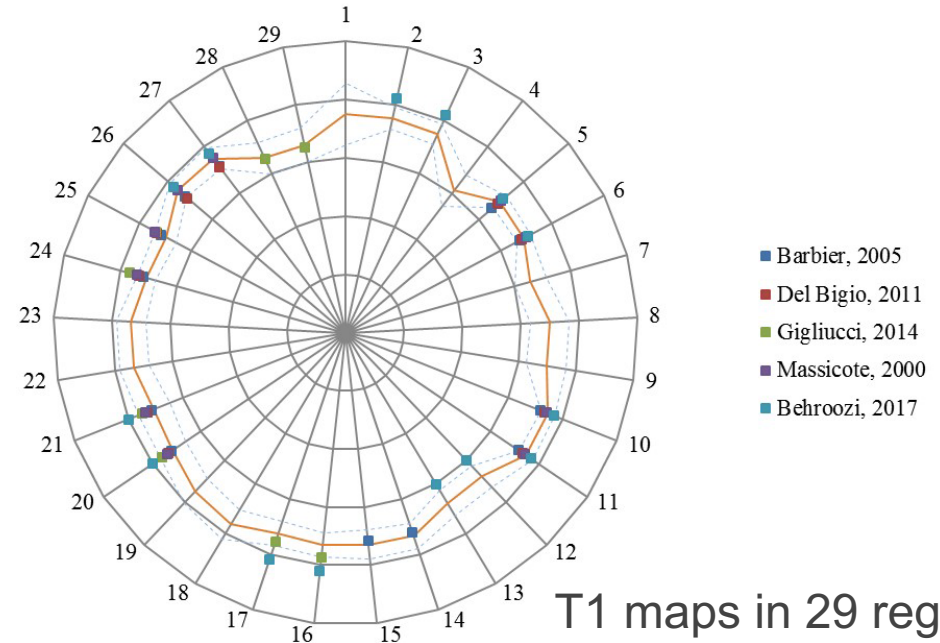
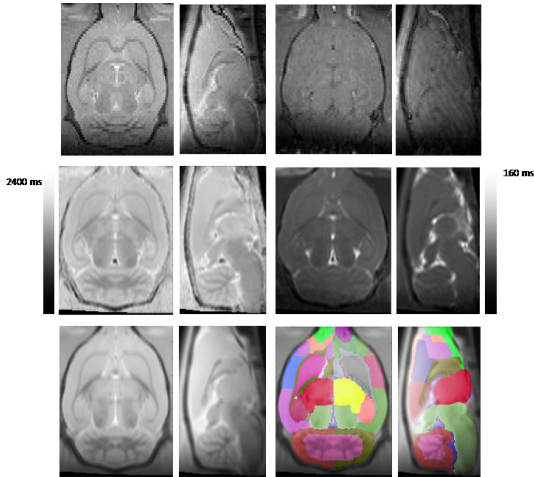
- Dicom + Json converter
- Ontoneurolog extension



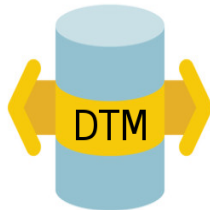
# FLI-IAM: Relaxometry Reference Maps



2 data provider centers  
(20 rats x2) 7T  
3 fitting pipelines  
2 multi-atlas pipelines



T1 maps in 29 regions



Deruelle et al submitted



# FLI-IAM: Relaxometry Reference Maps



6 data provider centers (60 mice)  
4.7 T, 7T, 9.4 T, 11.7T, 17.2 T  
2 multi-atlas pipelines





# FLI-IAM: For Preclinical Studies

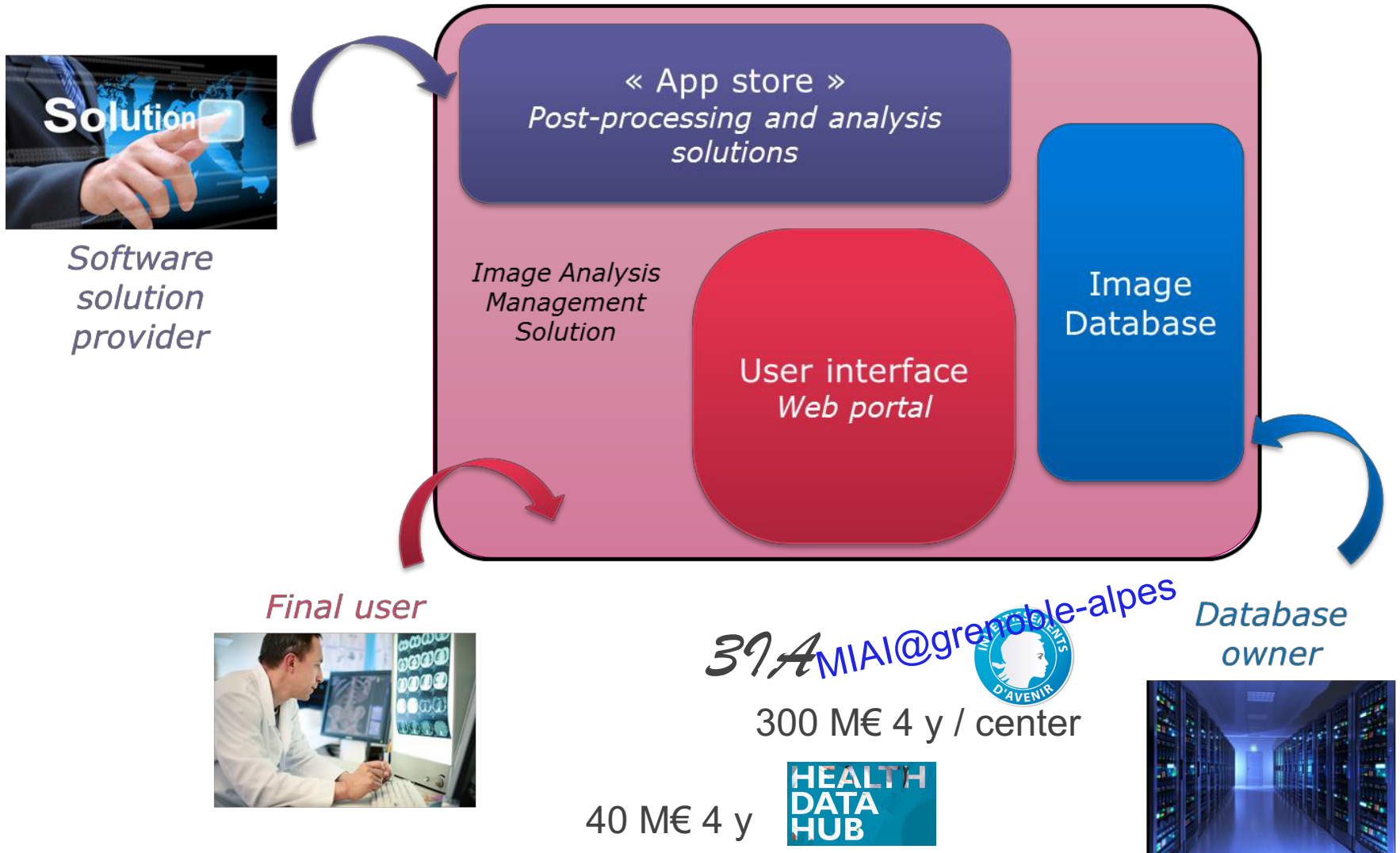
- Provide the control over sharing your data
- Manages your study meta-data and preclinical images
- Pathology models, therapies, anaesthetics and physiological data
- Import **Bruker** and DICOM file formats
- Secures online data sharing and data reuse
- Provides 10 year storage in the cloud
- Processes your preclinical images on high performance systems
- Supports you to integrate your data analysis pipeline
- Enriches DOIs to Open Access, e.g. OpenAIRE

# FLI-IAM: Major points

---

- Integration of large data sets for population-wide studies and the construction of imaging cohorts
- Validation and quality control of image processing tools on reference datasets
- Reuse and sharing of image processing pipelines on different datasets to improve research productivity and reproducibility
- Validation of research results based on the proofed control statistical analysis of images for quality control of experimental research

# FLI-IAM: Infrastructure Deployment + Services offer



# FLI-IAM: A collaborative action

## Scientific Committee



## Working Groups (Interoperability, Image processing pipelines, Preclinical imaging)





[contact-fli-iam@inria.fr](mailto:contact-fli-iam@inria.fr)

