



# **Archivage et partage des données d'imagerie : en quoi le big data modifie- t-il les pratiques?**

**C. Barillot, *DR CNRS***

*Inserm U746*

*IRISA (UMR CNRS 6074), Rennes*

*Centre Inria Rennes*

# Imagerie Médicale: un domaine du Big Data

- Bases de données en imagerie médicale
  - Entre 20% et 40% d'augmentation du stockage/an
  - 20% des actes répétés « inutilement »
  - Un hôpital moyen va produire entre 100-300 Tb d'images /an
  - Aux USA
    - 30% des images sont produites pour des raisons légales
    - Les images font parties de 400 millions d'actes médicaux / an
    - 1.3 milliards d'images médicales ont été stockées en 2010 (*~5M en 2015*)
  - Besoin croissant de stockage long-terme (recommandation EU de 20 à 70 an)
- Besoin en analyse (« *data analytics* ») : Traitement d'images, « machine learning »
- Requiert de la protection des données: contraintes légales complexes
- Très différent du domaine du e-commerce!: moins d'instances, plus de données / personnes

# Imagerie Médicale: un domaine du Big Data

- Bases de données en imagerie médicale
  - Entre 20% et 40% d'augmentation du stockage/an
  - 20% des actes répétés « inutilement »
  - Un hôpital moyen va produire entre 100-300 Tb d'images /an
  - Aux USA

**Besoin d'une adaptation technologique  
des solutions informatique pour la gestion  
des images médicales**

- ... contraintes légales complexes
- Très différent du domaine du e-commerce!: moins d'instances, plus de données / personnes

# Big Data et Imagerie Médicale: Etat de la situation

- Que faire des données produites?
  - Explosion des moyens de production et d'échange de données d'imagerie
  - Mais ... « *Information* » ne veut pas dire « *Connaissance* »
- Enjeux:
  - Comment exploiter cette masse d'information de manière aisée?
  - Comment traiter les images en masse ?
- Pour l'instant:
  - Approche plutôt basée sur une analyse descriptive que statistique, la recherche de corrélations ou encore sur la pensée que la masse compense la qualité

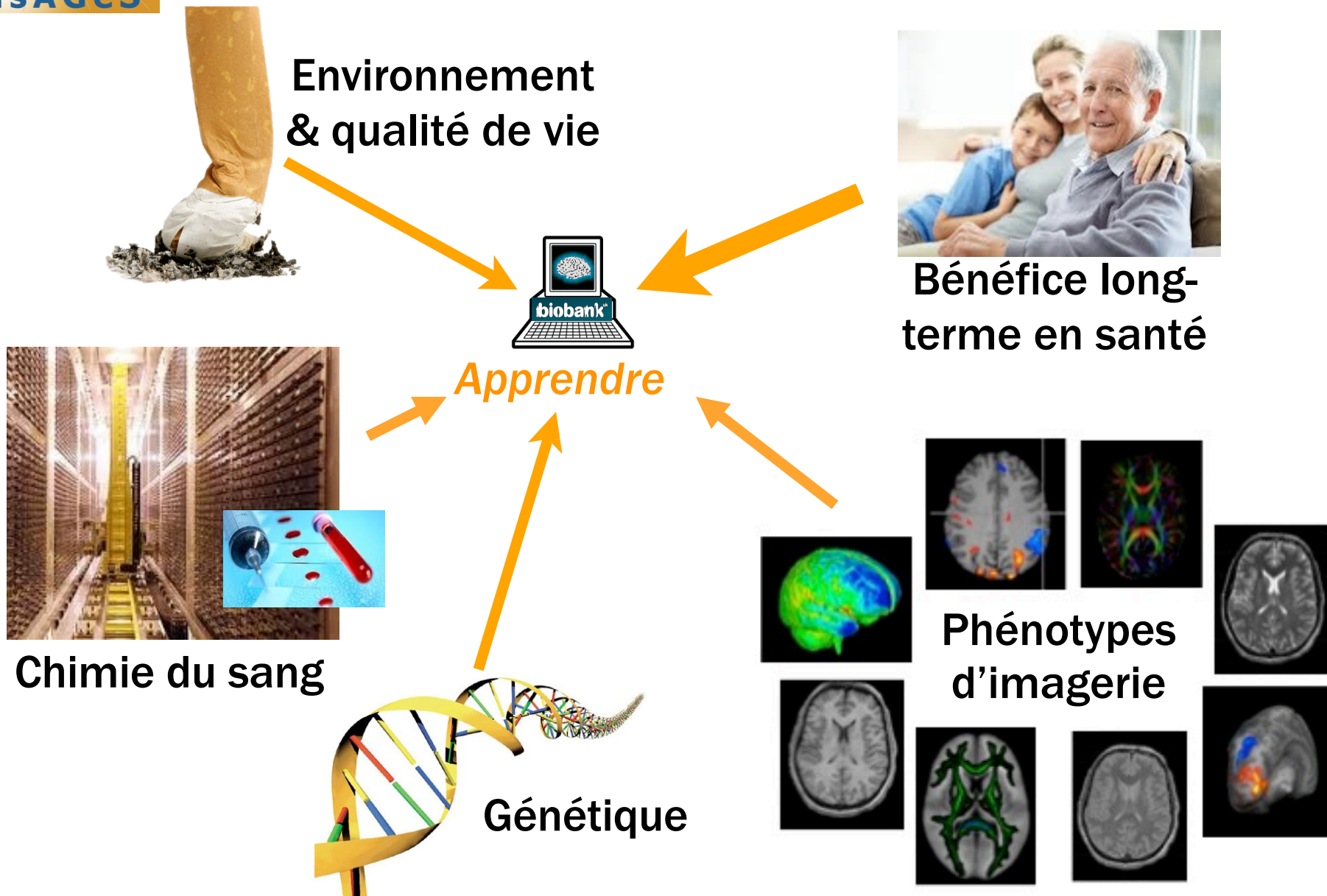
# Big Data et Imagerie Médicale: Vers où l'on va?

- Généralisation des infrastructures numériques et de l'Internet
  - Vers des « PACS 3.0 »
  - Stockage local dépassé
  - Dissociation acquisition/stockage
  - Console de visualisation et de traitement délocalisées
- Emergence d'infrastructures numériques dédiées
  - Quels opérateurs?
  - Quels couts et modèles de couts ?
  - Emergence de communautés virtuelles d'utilisateurs
  - Emergence de nouveaux services autour de l'image
- Emergence de nouveaux usages
  - Nouvelles façons de travailler: l'image se partage
  - Emergence de réseaux de soins virtuels
  - Evolution de la notion de couverture territoriale (l'image va au plus proche de l'expert): vers des centres d'imagerie spécialisés?
  - Partage et traitement des images → **Normalisation des protocoles**
  - Potentielles économies d'échelle

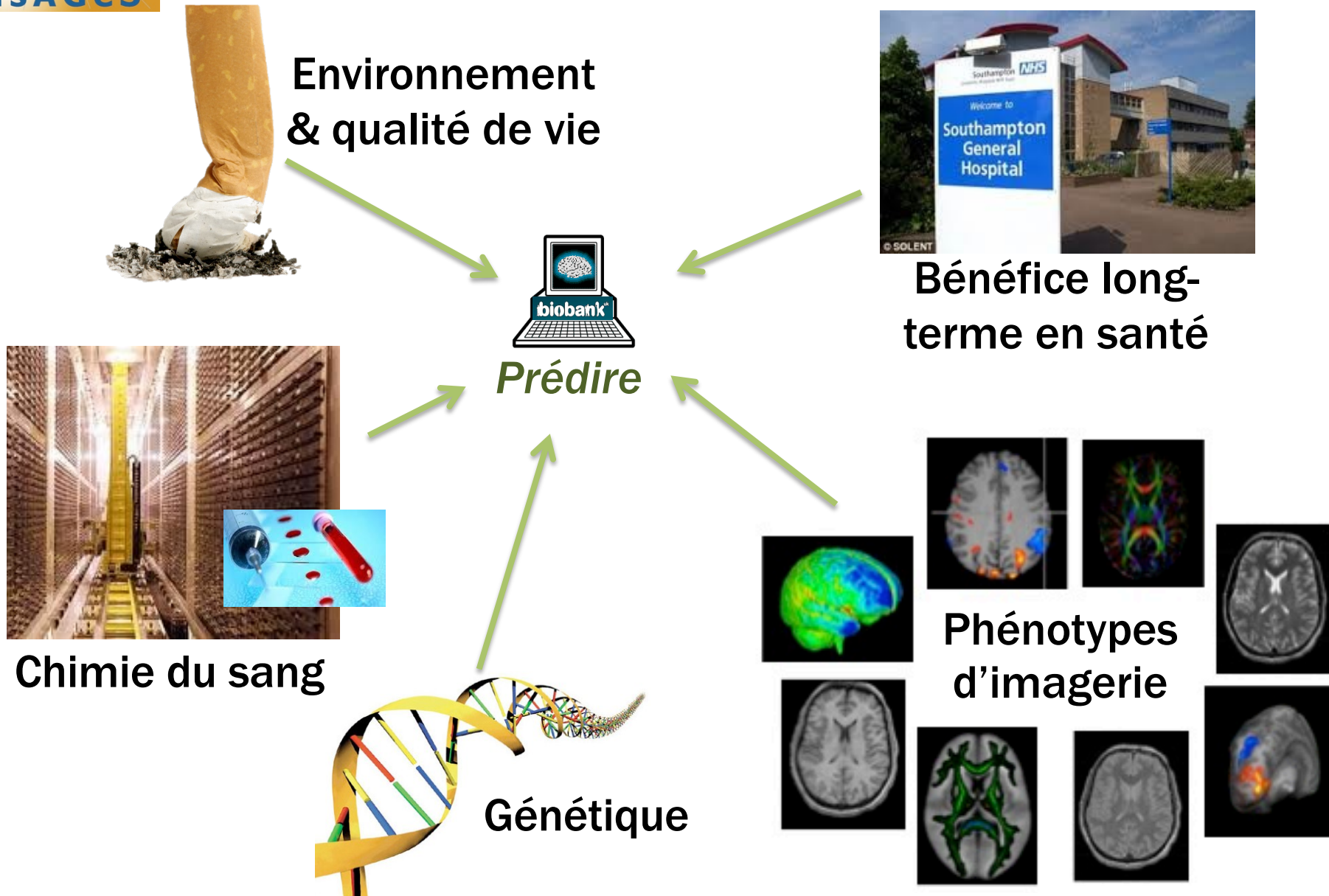
# Big Data et Imagerie Médicale: Vers l'Open Data?

- Besoin d'accéder à des jeux de données de plus en plus larges
  - Constitution d'atlas, apprentissage de modèles
  - Fouille de données, recherche de cas similaires
  - Enseignement / formation continue des personnels de santé
  - Evaluation / validation des outils de traitement d'images
  - Certification de solutions digitales autour de l'image
  - Incitation des agences de financement pour l'Open Data
  - **Emergence de l'imagerie de population**
- Enjeux
  - Quels opérateurs? (Publics, Privés / Nationaux, Mondiaux)
  - Quel modèle économique (qui supporte le cout?)
  - Quels Standards?
  - Contrôle qualité des données
  - Emergence de nouveaux acteurs (opérateurs réseaux, data centers, startup, GAFAM, ...)
  - Adaptation de la réglementation
  - Aspects éthiques

# Neuroimagerie en population: enjeux de santé



# Neuroimagerie en population: enjeux de santé



# L'imagerie médicale comme Service

- L'imagerie Médicale a des spécificités dans le monde du « cloud » :

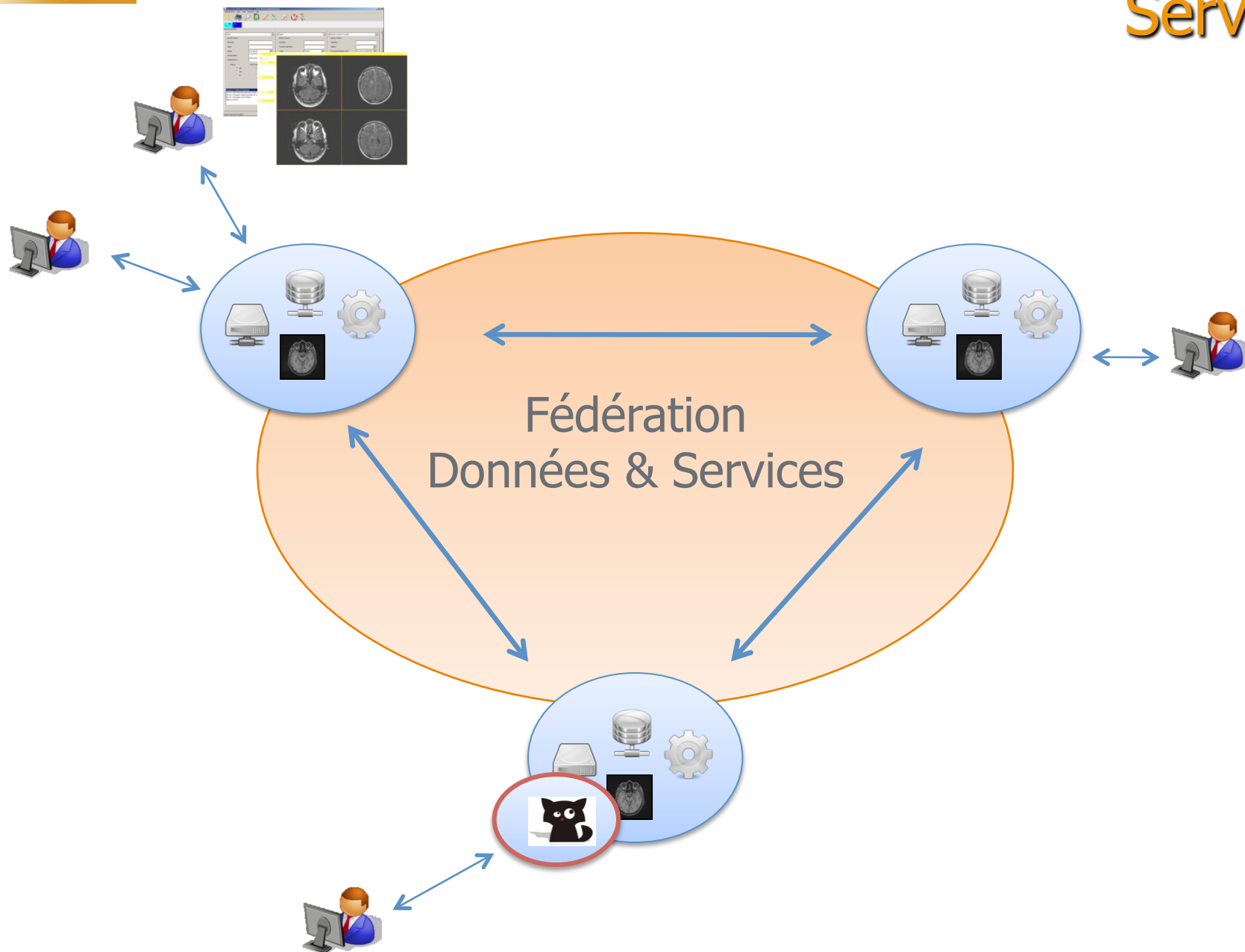
- Les Données:

- ◆ Sont hétérogènes
- ◆ Ont une sémantique forte
- ◆ Ont besoin de standardisation de protocoles d'acquisition
- ◆ Sont de plus en plus distribuées sur le territoire
- ◆ Sont confidentielles et doivent être sécurisées
- ◆ Doivent être conservées sur une très longue période

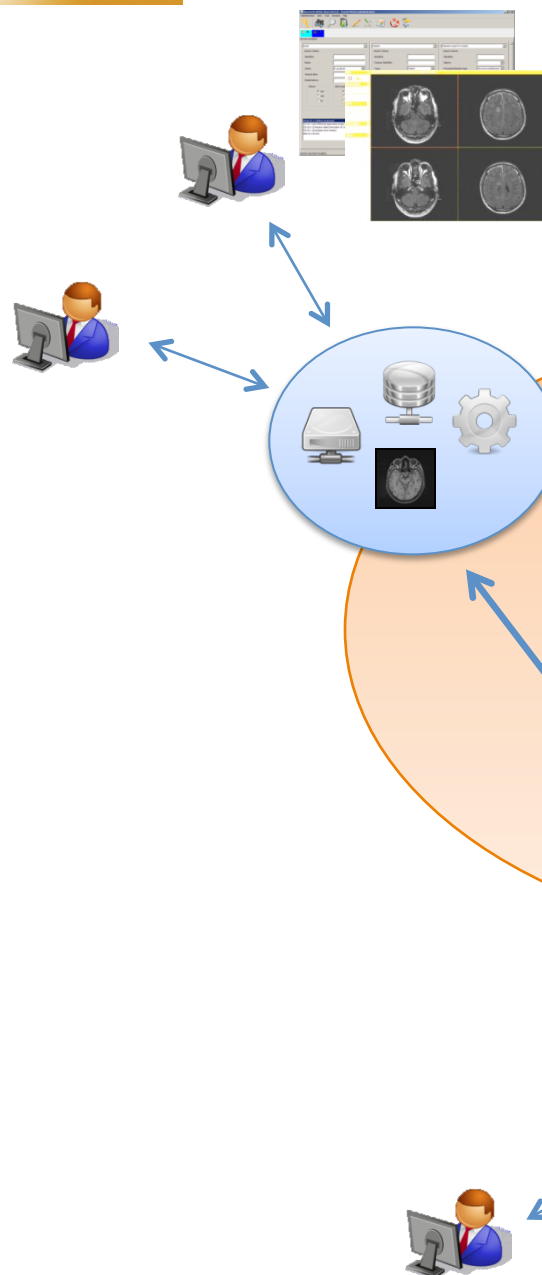
- Le traitement des données (Data Analytics):

- ◆ Sont souvent corrélés (workflows)
- ◆ Ont besoin d'automatisme pour les grandes cohortes (robustesse, adaptabilité, ..)
- ◆ Besoin de contrôle qualité sur les outils et leurs résultats: Evolution du qualitatif vers le quantitatif (biomarqueurs d'imagerie)
- ◆ Les temps de calcul peuvent être très importants sur des grandes cohortes
- ◆ ... mais les temps de calcul peuvent aussi être critiques (e.g. simulation temps-réel, interventionnel, urgence, ...)

# L'imagerie Médicale comme un Service



# L'imagerie Médicale comme un Service



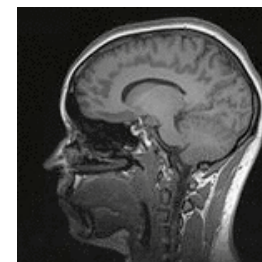
Un exemple de plateforme de gestion des images: Shanoir

Deux exemples of d'infrastructures de service émergentes :

- Cohorte nationale OFSEP
- Infrastructure France-Life-Imaging

# Motivations Pour Shanoir

- Imagerie: masse de données → toujours en augmentation
  - Modalités: MRI, PET, Scanner, ...
  - Exemples:
    - examen clinique standard IRM : ~ 1400 images (> 100Mo)
    - examen recherche standard IRM : ~ 4500 images (> 2Go)
  - Recherche Clinique → Etudes multicentriques, études multidisciplinaires



- Besoin de solution d'archivage
- Besoin de structuration des données
- Besoin de partager données et accès

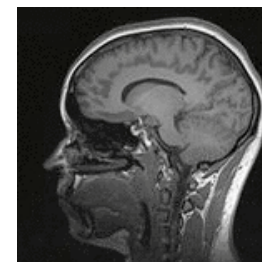


- Solutions existantes:
  - CDs: solution simple mais
    - Temps de vie imité (~ 5 ans), **Sécurité** et **Intégrité** des données non garanties
    - Problème de stockage physique
  - PACS: Picture Archiving and Communication System
    - Locaux (hôpitaux), réseaux régionaux → accès limités, solution mono-disciplinaire (radiologie)



# Motivations Pour Shanoir

- Imagerie: masse de données → toujours en augmentation
  - Modalités: MRI, PET, Scanner, ...
  - Exemples: - examen clinique standard IRM : ~ 1400 images (> 100Mo)



Une solution sécurisée de gestion des données d'imagerie dans le "cloud", accessible d'un simple navigateur Web

2Go) s multidisciplinaires



Il y a tellement plus facile pour partager ses données...



- Temps de vie limité (~ 5 ans), **Sécurité** et **Intégrité** des données non garanties
- Problème de stockage physique
- PACS: Picture Archiving and Communication System
  - Locaux (hôpitaux), réseaux régionaux → accès limités, solution mono-disciplinaire (radiologie)

## Qu'est ce que Shanoir ?

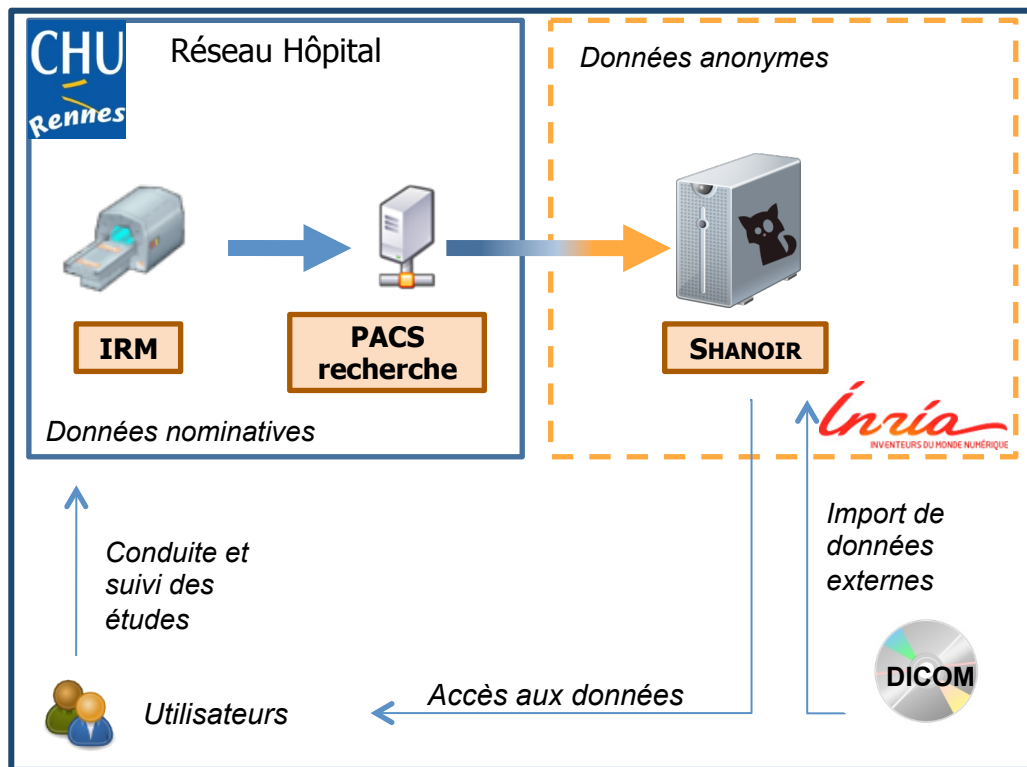
- Une solution **web** avec interface graphique **conviviale** et **sécurisée** permettant de :

- Archiver et Indexer
  - Chercher et Récupérer
  - Partager
- } des données d'imagerie



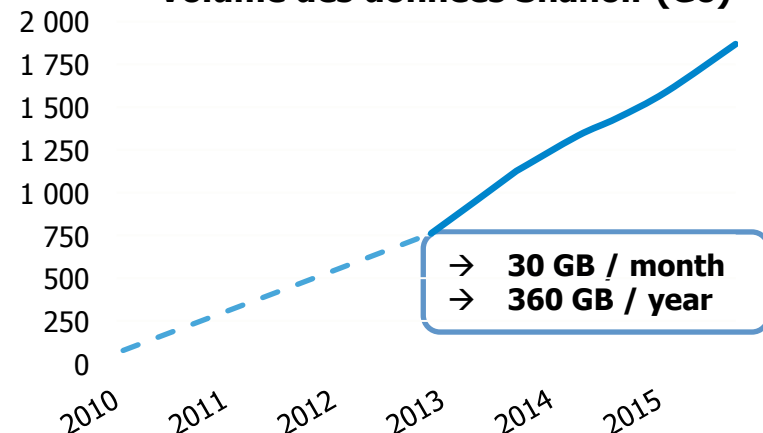
- Dans le but
  - D'améliorer l'**accessibilité** et l'**intégrité** des données
  - **Structurer** les données et gérer leur provenance
  - Faciliter la **collaboration** dans les études multicentriques
  - **Partager** les ressources d'acquisition

## Exemple : Déploiement de Shanoir dans le contexte de la plateforme Neurinfo de Rennes



	09/2013	09/2014	09/2015	05/2016
Centres	27	31	42	50
Users (active)	(53)	118 (52)	178 (66)	206 (99)
Etudes	53	60	70	121
Sujets	1706	2228	2833	3243
Examens	2268	3157	4005	4531
Datasets	-	114 441	151582	176356
Raw data (Dicom)	1 126 GB (1.1 TB)	1 434 GB (1.4 TB)	1 588 GB (1.5 TB)	2060 GB (2 TB)
Données traitées (Nifti)	989 GB (0.9 TB)	1 331 GB (1.3 TB)	1 758 GB (1.7 TB)	1841 GB (1.8 TB)

### Volume des données Shanoir (Go)





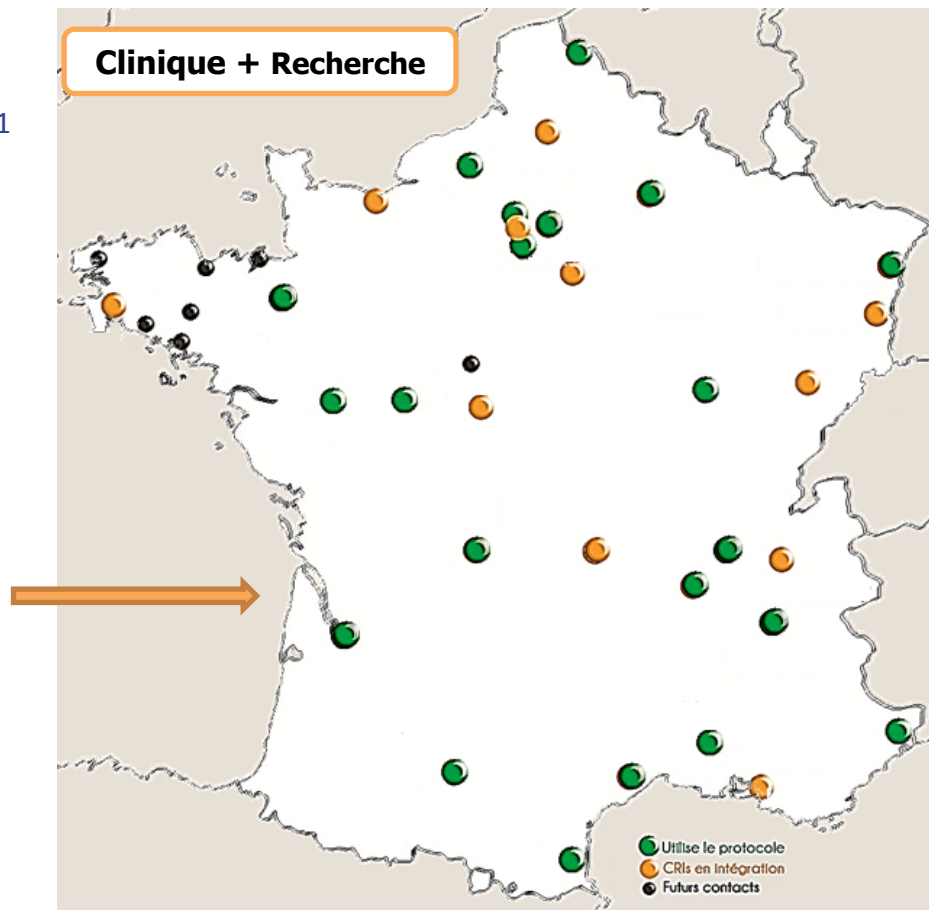
## Observatoire Français de la Sclérose en Plaques (PIA1-cohortes)

### Sclérose en plaques (SEP)

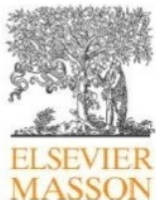
- 1,5 personne / 1,000 → 100,000 patients en France
- Construit autour d'un large réseau national:
  - 28 CHU (Depts. Neurologie et Radiologie)
  - 18 réseaux de soins régionaux
  - Historique basé sur le logiciel EDMUS, installé sur 61 sites

### Objectifs de la cohorte OFSEP

- 50,000 patients
- 1 IRM cérébral tous les 3 ans
- 1 IRM médullaire tous les 6 ans
- Pendant 10 ans  
→ **200,000** IRM avec un grand nombre d'équipements IRM
- ❖ ~ 300 sites d'acquisition
- ❖ Actuellement 14 différents modèles d'IRM
- Mise en place d'un protocole IRM standard au niveau national
- **Plateforme de gestion des données d'imagerie basée sur Shanoir**



Journal of Neuroradiology (2015) 42, 133–140



Available online at  
**ScienceDirect**  
[www.sciencedirect.com](http://www.sciencedirect.com)

Elsevier Masson France  
**EM|consulte**  
[www.em-consulte.com/en](http://www.em-consulte.com/en)



Cotton F, Edan G.  
L'Observatoire français de la  
sclérose en plaques : vers un  
consensus IRM national. **La  
Lettre du Neurologue**,  
juin 2015, n°6

## REVIEW

# OFSEP, a nationwide cohort of people with multiple sclerosis: Consensus minimal MRI protocol

F. Cotton<sup>a,\*,b</sup>, S. Kremer<sup>c</sup>, S. Hannoun<sup>b</sup>, S. Vukusic<sup>d</sup>,  
V. Dousset<sup>e,f,g</sup>, for the Imaging Working Group of the  
"Observatoire français de la sclérose en plaques" (OFSEP)

<sup>a</sup> Service de radiologie, hospices civils de Lyon, centre hospitalier Lyon Sud, Pierre-Bénite, France

<sup>b</sup> Inserm U1044, université de Lyon, université Lyon 1, CREATIS, CNRS UMR 5220, Villeurbanne, France

<sup>c</sup> Service d'imagerie 2, FMTS, université de Strasbourg, hôpital de Hautepierre, ICube UMR 7357, CHU de Strasbourg, Strasbourg, France

<sup>d</sup> Service de neurologie A et fondation Eugène-Devic EDMUS pour la sclérose en plaques, hôpital neurologique, hospices civils de Lyon, Bron, France

<sup>e</sup> Institut de bio-imagerie de Bordeaux UMS 3428, University of Bordeaux, Bordeaux, France

<sup>f</sup> Inserm, U862, physiopathologie de la plasticité neuronale, Bordeaux, France

<sup>g</sup> Service de neuroimagerie diagnostique de thérapeutique, CHU de Bordeaux, Bordeaux, France

Available online 7 February 2015

Cotton F, Kremer S, Hannoun S, Vukusic S, Dousset V; Imaging Working Group of the "Observatoire français de la sclérose en plaques" (OFSEP). OFSEP, a nationwide cohort of people with multiple sclerosis: Consensus minimal MRI protocol. **J Neuroradiol.** 2015;42(3):133-140.

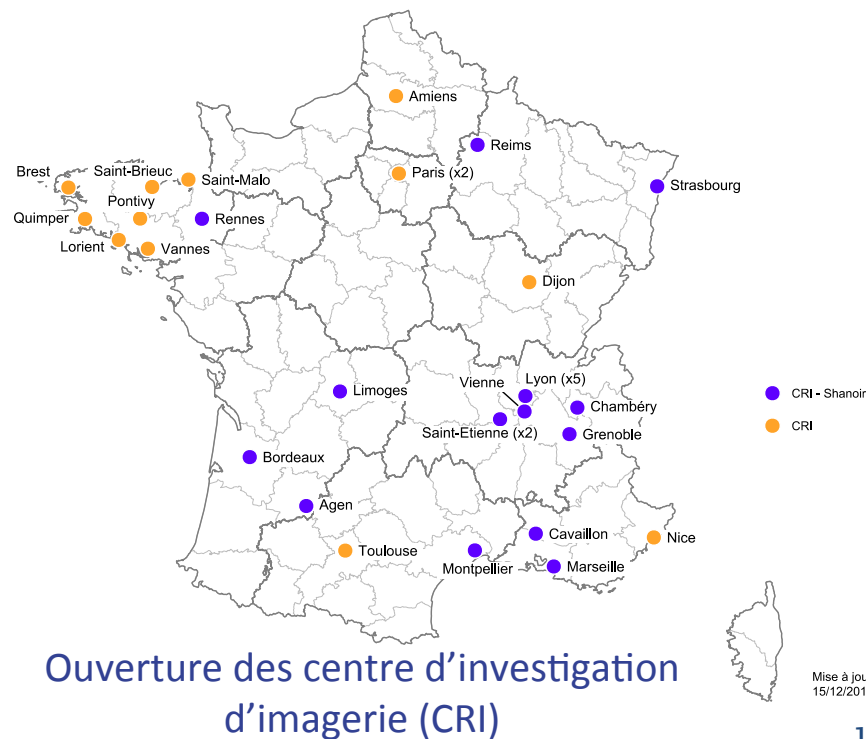
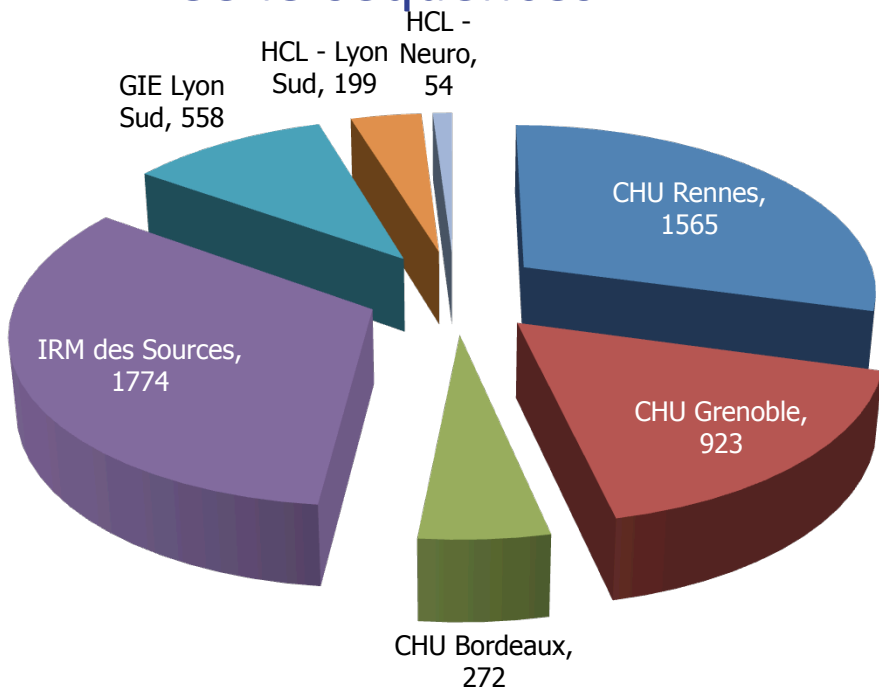


# Import des données IRM : Phase I (Phase Pilote)

- Examens stockés sur Shanoir (Décembre 2015) :

- 443 patients
- 553 examens
- 5345 séquences

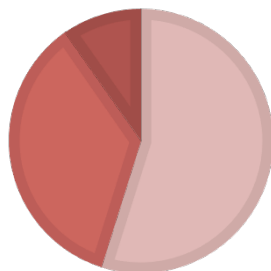
Phase II: Implémentation de l'accord CNIL puis développement de Shanoir Uploader



## Phase II: Variabilité des équipements inter-intra centres

	Constructeur / Modèle
8	Siemens - Aera 1,5T
7	Philips - Achieva 3T
2	General Electric - DISCOVERY MR750w 3T
2	Philips - Ingenia 1,5T
2	Siemens - Avanto 1,5T
2	Siemens - Skyra 3T
1	General Electric - Signa HDxt 3T
1	Philips - Achieva 1,5T
1	Philips - Ingenia 3T
1	Siemens - Espree 1,5T
1	Siemens - Spectra 3T
1	Siemens - Symphony Tim 1,5T
1	Siemens - Trio 3T
1	Siemens - Verio 3T
31	

■ Siemens  
■ Philips  
■ GE



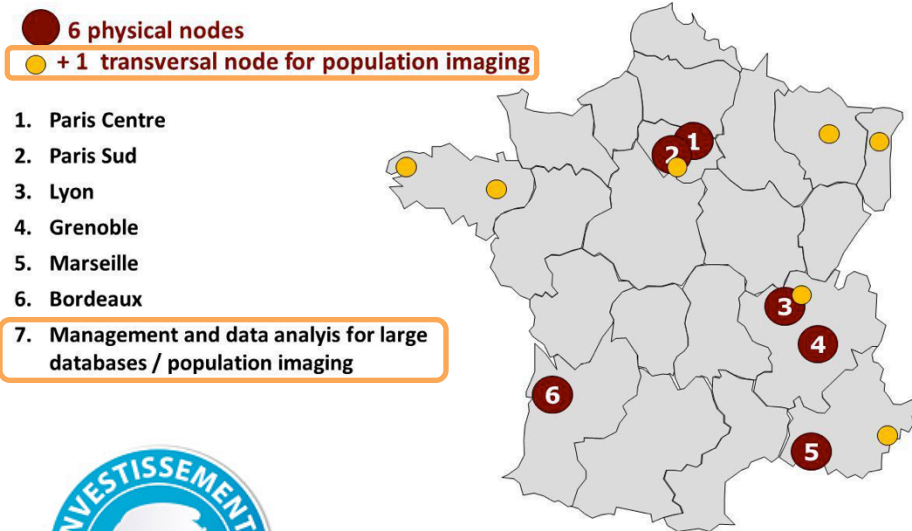
**Grande variabilité** dans chaque centre:

- Différents **équipements d'IRM**
- Différentes **versions logicielles** sur des équipements identiques
- Besoin d'installation du protocole sur les différentes machines
- Différents **modèles de PACS**
- **Différents usages dans les DSI**

	Constructeur	Modèle
4	GE Healthcare	Centricity PACS
3	Maincare Solutions	McKesson Radiology
2	Agfa	IMPAX
2	Telemed	Telemed
1	Carestream	Carestream Vue PACS
1	Global Imaging	GXD5 Pacs
13		

→ **Multiplication des contraintes de mise en place de l'infrastructure**

# Exemple d'une infrastructure large échelle : « France Life Imaging »



Nœud « Information  
Analysis and  
Management » (IAM)

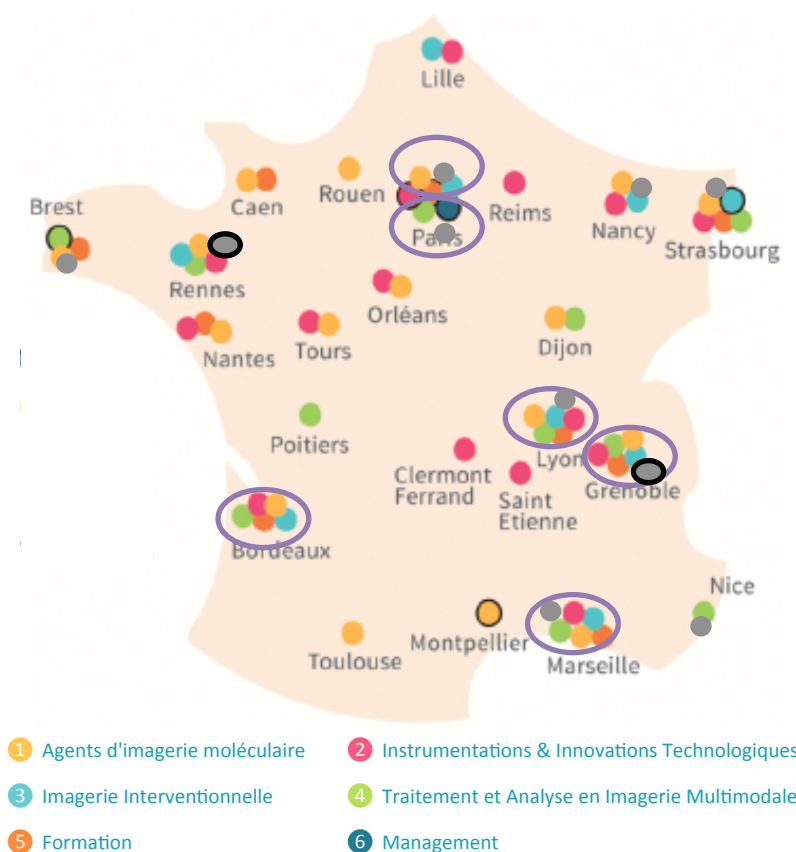


<https://www.francelifeimaging.fr>

<https://project.inria.fr/fli/>



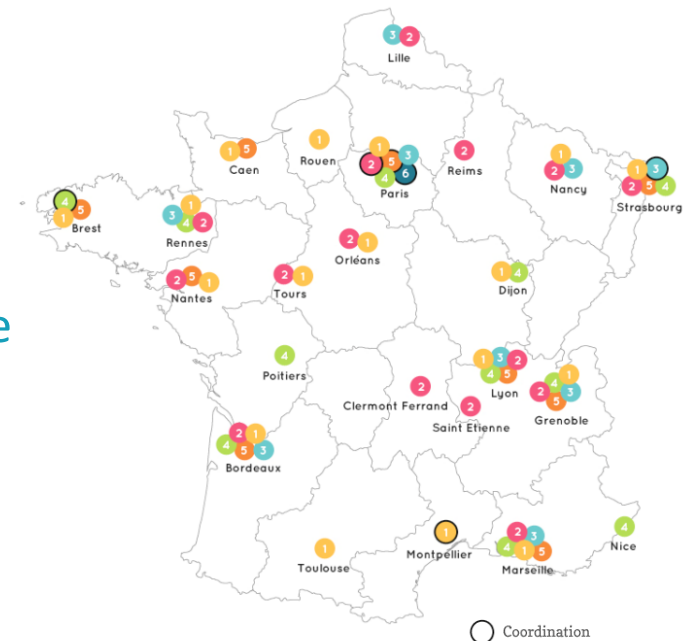
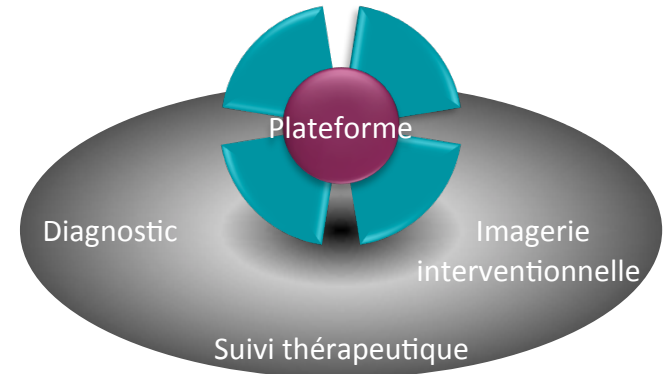
- Un réseau fédérateur couvrant tout le spectre de la recherche en imagerie *in vivo*
- Une approche translationnelle construite autour d'un lien structurel avec la clinique
- Un interlocuteur privilégié et reconnu pour la recherche partenariale
- Un point d'entrée unique pour trouver les interlocuteurs les mieux à même de répondre à vos besoins
- Créé en 2012 par le Programme Investissements d'Avenir (PIA2), afin de:
  - Doter la France d'équipements d'imagerie *in vivo* très innovants
  - **Créer des services de gestion et analyse des données**
  - Mettre à disposition de la communauté des chercheurs, des cliniciens et des industriels ces nouveaux équipements à travers un « guichet unique »



# FLI: Un réseau fédérateur couvrant tout le spectre de la recherche en imagerie *in vivo*



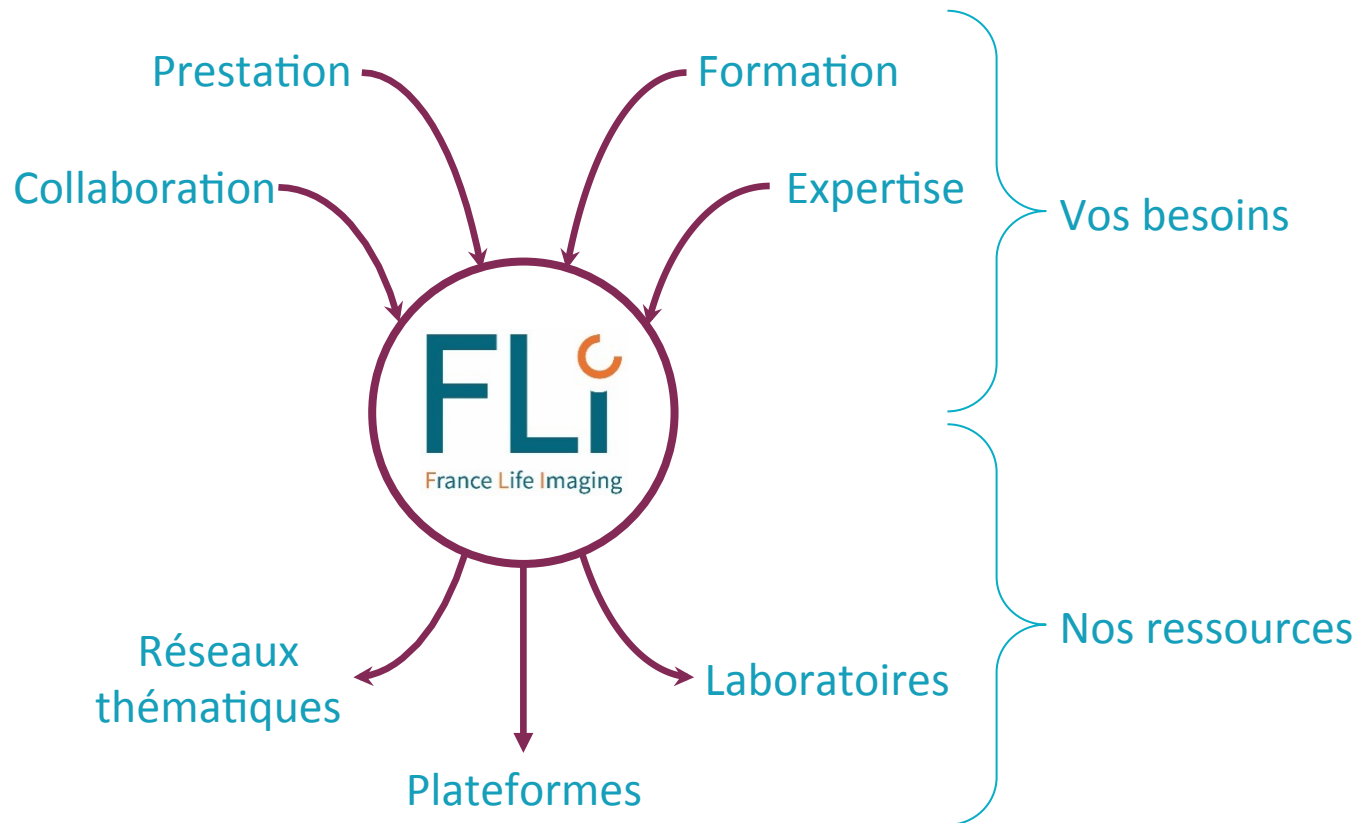
- Une organisation dédiée aux enjeux et besoins de la recherche en imagerie
- Une couverture exhaustive des champs thématiques majeurs du domaine
- Une animation de l'ensemble de la communauté à travers de nombreux événements et actions
- Des lieux de regroupement des équipements
- Des lieux de convergence des expertises scientifiques, médicales et technologiques
- Des lieux d'accueil de projets et d'équipes, de rencontre et d'échanges



○ Coordination



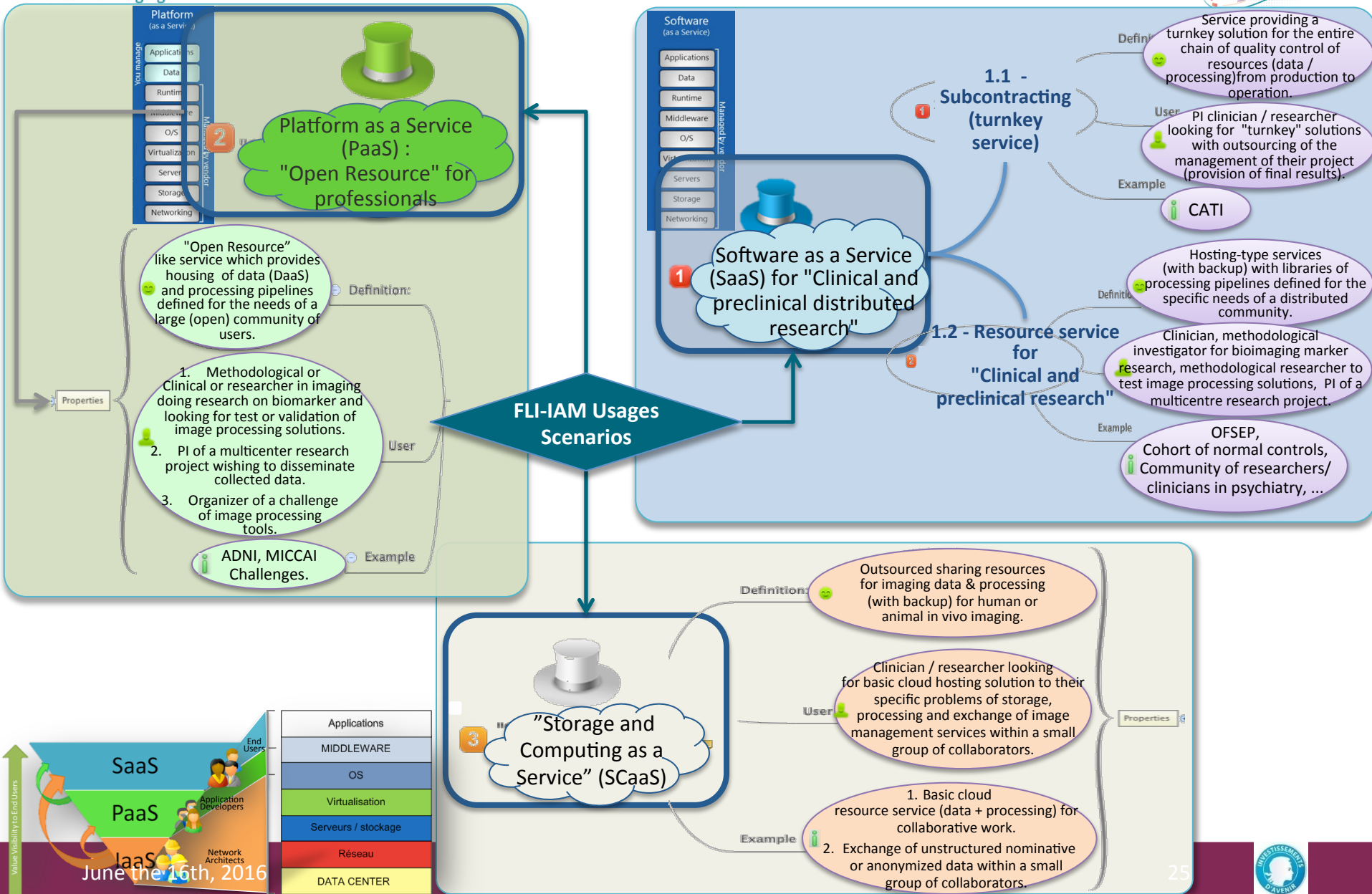
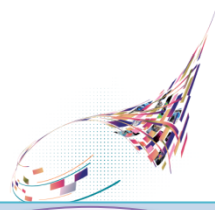
Un point d'entrée privilégié pour trouver les interlocuteurs les  
mieux à même de répondre à vos besoins

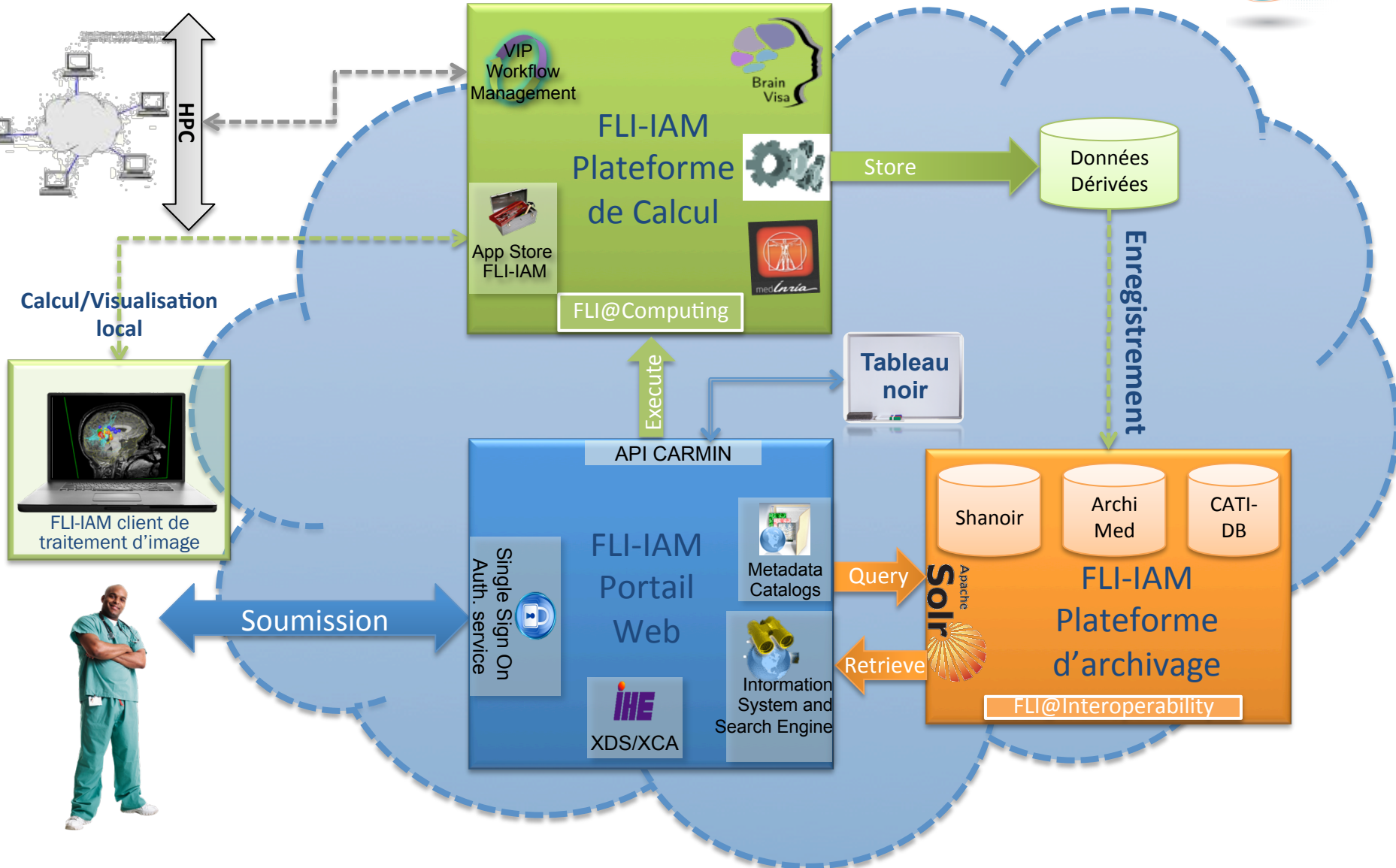




- Le nœud IAM est le nœud « STIC » de FLI
- FLI-IAM a pour ambition de construire et mettre en place une infrastructure informatique de gestion et traitement des données d'imagerie in-vivo en France, pour:
  - **Le clinicien** : pour conduire des projets de recherche clinique et préclinique impliquant l'imagerie médicale in-vivo
  - **La Pharma ou la start-up**: pour fournir des solutions digitales permettant de mieux gérer les images médicales ou pour fournir de nouveaux services au dessus des images
  - **Le chercheur en imagerie médicale**: pour permettre l'expérimentation et la validation de nouvelles solutions innovantes en imagerie médicale
- **Déploiement de FLI-IAM en deux phases**
  - **Phase I (2013-2017):**
    - Mise en place de l'infrastructure sur la base de solutions logicielles existantes (Shanoir, MedInria, CATI, VIP, Archimed, ...) et sur la base de l'analyse des futurs usages
    - Phase pilote basée sur un nombre limités de cas d'usages en imagerie médicale et imagerie préclinique
    - Préparation d'une offre de service
  - **Phase II (après 2017) :**
    - Phase Opérationnelle, ouverture du service
    - Arrivée d'opérateurs (publics/privés) pour la gestion opérationnelle de l'infrastructure et la fourniture des équipements
    - Extension de l'infrastructure pour accueillir de nouveaux usages et nouveaux services

# FLI-IAM SCENARIOS d'USAGES





# Archivage et partage des images: Comment changer les pratiques?

- Le regroupement des données reste un enjeux
  - Réconciliation des sources d'images, réconciliation des images avec leurs contexte clinique, fusion des images, contrôle qualité, harmonisation
- L'analyse des images reste un domaine en émergence
  - Robustesse des outils, temps de calcul, reproductibilité des résultats, responsabilité des usages
- Qui veut partager ses données?
  - Conservatisme de la communauté
  - Contraintes règlementaires
- Qui peut partager ses données?
  - Dépasser le stade du « club » des initiés
  - Offrir des solutions adaptées aux usages
  - Certification des solutions fournies
  - Accepter de payer pour ces nouveaux usages. (quel modèle économique pour ces usages?)
- Comment gérer éthiquement le partage des données et l'open data?
  - Anticiper avant que les problèmes n'arrivent : collectivement (sur le plan légal) et individuellement (e.g. consentement)
- La sécurité reste et restera toujours un enjeux
  - Ne pas la sous-estimer
  - Ne pas s'en servir d'excuse pour ne pas faire
  - Bien dimensionner la réponse au risque pour ne pas tuer l'usage
  - Multidisciplinarité: La sécurité informatique n'est pas qu'un enjeux en médecine
- Inventer de nouveaux métiers et les valoriser
  - Rien ne pourra se faire sans une intégration fortes entre Ingénieurs, chercheurs, juristes et médecins

