



**HAL**  
open science

## Imageurs, encore un effort si vous voulez partager vos données !

Michel Dojat

### ► To cite this version:

Michel Dojat. Imageurs, encore un effort si vous voulez partager vos données!. Ouvrage collectif sur les nouvelles imageries pour la santé, A paraître. hal-03273927

**HAL Id: hal-03273927**

**<https://hal.science/hal-03273927v1>**

Submitted on 29 Jun 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## **Imageurs, encore un effort si vous voulez partager vos données !**

**Michel Dojat**

Au-delà d'une contribution scientifique originale, cet article est avant tout un témoignage. Il rend compte du compagnonnage intellectuel qui nous a rassemblés, avec Christian Barillot, durant presque une vingtaine d'années. Il montre comment autour d'une idée initialement peu élaborée, se construit en commun une réflexion scientifique mature. Comment cette activité créatrice progresse en fonction des collaborateurs qui vont y être associés et des supports financiers que l'on va obtenir. Cet article raconte une histoire, une histoire de *partage*, notion qui convient bien à la personnalité de Christian Barillot qui, tout au long de sa carrière de chercheur et de directeur d'équipe, a voulu partager avec d'autres sa passion pour les questions scientifiques soulevées par l'analyse de données d'imagerie médicale. C'est donc plutôt une contribution à l'étude de « la vie de laboratoire » ; non pas conduite par un observateur en immersion dans un environnement scientifique (comme peut la décrire Bruno Latour dans son ouvrage *La vie de laboratoire*, 1988), mais, fait plus rare, incarnée et rapportée par un des acteurs de la production de ces « faits scientifiques ».

Le titre de cet article fait écho à celui d'un chapitre politique de « La philosophie dans le boudoir » du Marquis de Sade : « Français, encore un effort si vous voulez être républicains » (1795) qui, au-delà du clin d'œil qui aurait sans doute amusé Christian, d'un rapprochement improbable entre dialogue libertin et révolutionnaire et rhétorique scientifique, veut souligner la posture engagée de Christian Barillot pour la défense du service public, son exigence de rigueur d'un travail orienté vers la communauté, et indiquer que finalement *partager* est aussi une posture avant tout humaniste et politique.

### **Pourquoi partager des données d'imagerie**

Pour confirmer ou infirmer une hypothèse, les scientifiques font des expériences spécifiques. Ils doivent ensuite rapporter leurs observations et leurs résultats d'analyse très précisément avec les conditions sous lesquelles ils ont été obtenus (banc expérimental, méthodes de mesure, algorithmes et procédures de traitement des données etc.), dans le but de les rendre reproductibles. En effet, pour qu'une observation ou un résultat se transforme en fait scientifique, il faut qu'après différentes études de reproduction, le cas échéant en contrôlant les biais relevés par les chercheurs, il soit possible, au moins provisoirement, de l'accepter pour vrai. Un premier intérêt du partage de données est de pouvoir les analyser avec une procédure différente de celle initialement utilisée. Ceci est particulièrement vrai dans le domaine de l'analyse d'images où de nouveaux algorithmes, des capacités informatiques plus performantes, peuvent permettre de détecter des différences jusqu'alors indétectables entre un groupe de sujets sains et un groupe de patients. En science de la vie, les études

d'imagerie *in vivo* sont lourdes et coûteuses à réaliser. Il est donc intéressant, notamment pour établir des normes physiologiques sur des sujets sains et étudier ensuite les écarts dans les cas pathologiques, de pouvoir rassembler des données obtenues dans différents centres. Par ailleurs, le nombre de sujets que l'on peut étudier dans un seul centre est généralement faible (soit une trentaine maximum, sur une durée inférieure à un an). Ainsi, une analyse récente d'études incorporant de la neuroimagerie fonctionnelle par IRM (Imagerie par Résonance Magnétique) montre que, même si la taille des échantillons est en constante progression depuis 1995, date de son introduction comme méthode non invasive d'étude du fonctionnement cérébral, cette taille reste faible, 28 sujets en moyenne. Cela reste insuffisant pour mettre en évidence des différences de valeur modeste ; ce qui est le cas de la majorité des études. Cette faible puissance statistique rencontrée dans beaucoup d'études en sciences de la vie conduit à la publication de résultats non reproductibles et à la production de faux positifs (la différence n'existe pas et on la trouve) et de faux négatifs (la différence existe et on ne la trouve pas). Le partage de données entre centres de recherche est donc un des éléments pour améliorer la robustesse statistique des résultats. L'apparition de la technique non invasive d'IRM fonctionnelle entraînant la mise à disposition d'une quantité de données importante sur le cerveau, a incité les neuroscientifiques à une réflexion sur l'organisation et le partage de leurs données plus précocement que dans d'autres domaines des sciences de la vie. Récemment, le [COBIDAS](#) (Committee on Best Practices in Data Analysis and Sharing) a publié les bonnes pratiques pour l'analyse des données et le partage en Neuroimagerie. Nous avons démarré avec Christian Barillot une réflexion sur ce thème il y a une vingtaine d'années.

### **Aux sources du partage de données**

Le projet NeuroBase (2002-2004) coordonné par Christian Barillot, a rassemblé des équipes multidisciplinaires en santé, traitement d'images et informatique autour d'un objectif commun : la définition d'une architecture informatique pour connecter différentes bases de données d'imagerie *in vivo*. Une approche plus standard utilisée dans de gros projets nord-américains (ADNI, MJ Fox...) consiste à rassembler toutes les données dans une base de données unique avec un schéma d'organisation prédéfini. L'architecture fédérée proposée dans NeuroBase offre plus de flexibilité permettant l'accès et le partage de résultats d'expérience stockés dans des dépôts distribués préexistants. Elle peut aussi faciliter l'accès à des procédures de traitement de données réparties dans différents centres et qui souvent nécessite l'installation d'environnement d'exécution spécifique (système d'exploitation ou bibliothèques de logiciels). NeuroBase a mis en évidence la nécessité de développer un modèle sémantique de référence, modèle « public », vers lequel est mis en relation (mapping) chaque modèle « privé » d'organisation des données, spécifique à chaque base de données fédérée. L'utilisateur a alors une vue globale sur les données disponibles. Il pose des requêtes via le

modèle public (par exemple : « recherche images anatomiques et de perfusion cérébrales pour des patients Parkinsoniens d'âge inférieur à 65 ans avec un score de gravité sur l'échelle de Hoehn et Yahr inférieur à 3 ») et il accède aux images répondant aux critères et disponibles sur l'ensemble des bases ; et ceci indépendamment de la façon dont elles sont sémantiquement organisées localement. Ce modèle de référence se devait de couvrir, pour un large corpus de requêtes, les différents types d'images rencontrées (IRM, tomographie par émission de positrons, imagerie par rayons X, etc.), les procédures utilisées pour les acquérir et les traiter, les conditions d'expérimentation (comme les paradigmes de stimulation utilisés en imagerie fonctionnelle), les résultats d'analyse et les annotations qui pouvaient être reportées sur les images.

Pour construire un tel modèle de référence public, nous avons identifié l'ensemble des concepts présents dans les bases de données candidates et construit une « ontologie d'application », soit une hiérarchie de concepts avec les différentes relations qui peuvent les relier. Dans un second temps nous avons utilisé un éditeur d'ontologie, « Protégé », pour transcrire informatiquement cette ontologie. La figure 1 montre une partie de cette ontologie (figure 1, gauche) et les liens qui doivent être faits (mapping) pour la relier à une organisation locale (figure 1 droite) et spécifique des données. Ainsi, le concept public de « sujet » participant à une étude est relié dans l'organisation locale des données à « subj » et celui de « Dataset » aux différents types de données organisées dans des sous-répertoires. Ensuite, pour opérationnaliser le « mapping » des référentiels privés vers le référentiel public, nous avons opté pour une approche de type « médiateur/ adaptateur ». Pour démontrer la pertinence d'une approche fédérée de sources d'information distribuées et hétérogènes, nous avons réalisé une expérience permettant la circulation de l'information entre des données acquises par un centre, l'unité de Neuroimagerie fonctionnelle et métabolique de Grenoble, partagées et prétraitées par les outils d'un second centre, l'équipe Vista de l'IRISA de Rennes, puis segmentées (repérage automatique des tissus) de façon concurrente par les outils des deux partenaires de façon à comparer les performances. Les résultats étaient visualisés grâce à l'outil provenant de l'institut fédératif de recherche de neuroimagerie fonctionnelle.

Fort de cette première réalisation nous avons obtenu un financement de quatre ans pour le projet NeuroLOG (2007-2010). Aux côtés des neuroimageurs de NeuroBase, NeuroLOG se renforce alors sur l'axe informatique avec, pour la composition et l'exécution de pipelines de traitements d'images, des spécialistes du calcul distribué (laboratoire de recherche informatique, LRI, de Saclay et laboratoire d'informatique, signaux et systèmes, I3S, de Sophia-Antipolis), pour le raffinement du modèle public de données, des ontologues (laboratoire de recherche en informatique d'Amiens) et pour disposer de briques logiciels robustes, des industriels en génie logiciel (entreprise Business Objects) et en outils de visualisation d'images 3D (entreprise Visioscopie). Ces nouvelles forces nous ont permis de raffiner le

module de requête, le modèle sémantique et l'architecture logicielle. Ainsi, le modèle initial de données a été reconstruit pour reposer sur une ontologie générique (DOLCE, pour *Descriptive Ontology for Linguistic and Cognitive Engineering*), qui offre un haut niveau d'abstraction facilitant l'interopérabilité avec d'autres ontologies d'application. De plus, il a été étendu pour introduire les batteries de tests de la neuropsychologie au côté des données d'imagerie. Cela permet de disposer des scores de comportement et de divers résultats d'examens cliniques neurologiques et cognitifs. Cette ontologie inclut une taxonomie générale des instruments permettant d'accéder aux capacités cognitives et mémorielles, tels que l'échelle de statut mental (MMS) ou « de Hoehn et Yahr », et donc de faire des requêtes intégrant ces informations. Dénommée [OntoNeuroLOG](#), cette ontologie d'application est disponible via Bioportal, le portail des ontologies biomédicales. Elle est aussi au cœur du développement de l'outil de gestion de données [Shanoir](#) utilisé dans de nombreuses applications de recherche clinique et préclinique en neuroimagerie *in vivo*.

La vue d'ensemble de l'architecture logicielle est présentée sur la figure 2. L'interface orientée Internet (niveau I) permet à l'utilisateur du système de formuler des requêtes de façon simple et conviviale sur des images médicales, des traitements et de visualiser le résultat. Ces requêtes sont converties (niveau II) dans un langage informatique (RDF, pour Resource Description Framework), puis exécutées *via* l'outil *Corese* pour permettre une recherche sémantiquement enrichie grâce à l'ontologie au cœur du système, qui regroupe tous les termes et concepts qui constituent le champ d'intérêt. L'outil *METAmorphose* fait le lien entre le langage RDF (qui décrit les ressources Web et leurs métadonnées) et le langage standard (SQL, pour Structured Query Language) qui permet de communiquer avec les bases de données. L'outil *DataFederator* (niveau III) fait le lien entre les représentations locales des bases de données fédérées (NeuroLOG data base) et le modèle public intégré dans l'ontologie. Le module IV est utilisé pour composer (outil MOTEUR) graphiquement des pipelines de traitement d'images et permettre leur exécution sur les données images sélectionnées via des environnements tels que des grilles de calcul.

Le projet NeuroLOG a démontré la faisabilité et l'intérêt d'une approche fédérée basée sur l'utilisation d'une ontologie de domaine pour l'interrogation de dépôts de données et d'outils en neuroimagerie. À noter que l'ontologie peut être étendue en vue de l'intégration de nouveaux types de données. Un tel modèle formellement et sémantiquement bien défini facilite l'interopérabilité avec des dépôts d'autres types d'informations (comme des données d'imagerie *in vitro* ou de génétique).

## L'analyse et la gestion des données issues de l'imagerie

Le projet France Life Imaging ([FLI](#)) lancé en 2013 a pour mission de coordonner à l'échelle de la France les plateformes de recherche en imagerie *in vivo* chez l'Homme et l'animal. Initialement construit autour de six « hubs » fournisseurs de données (Paris, Paris-Sud, Lyon, Grenoble, Marseille, Bordeaux), il a récemment (2020) été étendu à 3 hubs supplémentaires : Grand-Ouest (Rennes-Nantes), Grand-Est (Nancy-Strasbourg), Occitanie (Toulouse-Montpellier) pour un meilleur maillage du territoire. Ces plateformes comprennent diverses modalités d'imagerie : IRM, optique, ultrasonore, nucléaire, etc. L'imagerie IRM, technique non invasive, polyvalente, qui permet notamment d'imager le cerveau et de recueillir alors des mesures anatomiques, fonctionnelles et de perfusion cérébrale, est utilisée dans 72% des études cliniques et précliniques menées sur le territoire. Pour les besoins de la recherche, de nombreuses études sont réalisées sur de petits animaux (rat, souris), qui constituent des modèles pour certaines pathologies (comme Parkinson, Huntington, ...). Au niveau national, les plateformes d'imagerie *in vivo* sont utilisées en majorité (68%) pour ce type d'études dites « précliniques ». Les études cliniques et précliniques sont encadrées par des comités d'éthiques qui s'assurent qu'elles sont réalisées dans le respect de la personne humaine et de l'animal, qu'elles sont pertinentes scientifiquement sur base de la littérature déjà existante dans le domaine d'étude et que la méthodologie d'acquisition et d'analyse des données est adaptée. Pour l'expérimentation animale, la règle des 3R est de rigueur : réduire le nombre d'animaux étudiés ; raffiner la méthodologie utilisée pour obtenir plus d'informations pertinentes dans de bonnes conditions pour le sujet ; et remplacer les modèles animaux notamment par des modèles *in vitro* ou mathématiques (*in silico*). France Life Imaging permet à la fois de coordonner la jouvence du matériel, d'assurer la complémentarité des investissements et de mettre en place des réseaux d'experts sur différents axes de l'imagerie *in vivo* : agents d'imagerie, imagerie interventionnelle, instrumentation et traitement des images.

Le neuvième hub, actuellement piloté à l'Institut des Neurosciences de Grenoble, « Information Analysis and Management » (IAM), concerne la gestion et l'analyse de données d'imagerie associées aux plateformes. Il s'agit de construire une infrastructure numérique pour le stockage et le traitement des données d'imagerie *in vivo* à destination de trois types d'utilisateurs. Cette plateforme intéresse d'abord le clinicien/scientifique qui conduit une large étude multicentrique en recherche clinique ou préclinique associant de l'imagerie *in vivo*. Il veut donc pouvoir récupérer les données des centres participants, définir et exécuter des algorithmes pour traiter cette masse de données (en neuroimagerie : calculer les volumes des tissus ou des lésions du cerveau, par exemple) et enfin rendre visibles ces données à la communauté scientifique dans un but de partage. Le deuxième public concerné est le traiteur académique d'images ou la start-up qui veut expérimenter une nouvelle solution algorithmique, la valider ou la mettre à disposition – voire mettre en place un défi pour

comparer différentes solutions face un problème spécifique (comme repérer automatiquement les nouvelles lésions apparues entre deux consultations neurologiques chez les patients souffrant de sclérose en plaques). Enfin, les départements de recherche pharmaceutique et les sociétés de recherche sous contrat sont également en demande d'une plateforme opérationnelle pour gérer les données associées à leurs essais cliniques. Sur la base des travaux antérieurs, *NeuroBase* et *NeuroLog*, nous avons opté, avec Christian Barillot, pour une approche fédérant les solutions existantes de gestion de données d'imagerie. Un réseau d'experts du domaine a été constitué, comprenant des chercheurs de différents laboratoires en traitement d'images cliniques et précliniques. Sous la direction d'un chef de projet informatique recruté spécifiquement, nous avons recruté des ingénieurs informaticiens de façon à réaliser un démonstrateur opérationnel permettant la fédération de trois bases de données existantes : Shanoir (Inria), Archimed (Inserm) et Cati-DB (Cea). L'exécution des pipelines était réalisée sur la plateforme VIP (CNRS). Par rapport aux travaux précédents, les forces en ressource humaine (quantifiées par l'unité « hommes-mois, h.m ») mises en jeu ont été plus conséquentes. Nous avons ainsi pu développer un portail web pour rendre visible données et outils (100 h.m), définir une interface spécifique (API Carmin) pour faciliter l'exécution de programmes de traitement sur VIP (98 h.m), développer un module permettant la récupération des résultats de traitement pour leur injection dans les bases de données sources (67 h.m) et enfin réaliser l'interopérabilité des modèles privés des trois bases de données fédérées vers le modèle public (220 h.m). L'architecture obtenue est représentée sur la figure 3. Le portail Internet permet à l'utilisateur de faire des requêtes sur les données disponibles et les chaînes de traitement à sa disposition. Plusieurs bases de données sont accessibles (module en orange, figure 3) selon le principe fédératif décrit précédemment. Sur les données sélectionnées, l'utilisateur peut exécuter sur la plateforme de calcul un pipeline disponible (module en vert figure 3). Un module (DTM, figure 3) assure l'échange d'informations pour stocker les résultats de traitement dans les bases de données correspondantes. Pour montrer l'intérêt de notre plateforme, nous avons organisé deux défis lors de la conférence MICCAI 2016 (*Medical Image Computing and Computer Assisted Intervention Society*). Le premier défi (Pet-Tumor) consistait à déterminer de façon automatique le volume tumoral dans des images par tomographie par émission de positrons. Ces images ont un faible rapport signal sur bruit et une faible résolution spatiale, ce qui rend la zone tumorale difficile à délimiter précisément. Le second défi (MSSEG) concernait la détection de lésions dans des données IRM cérébrales (images de type Flair) de patients souffrant de sclérose en plaques. Pour ces deux défis, les délimitations manuelles par des neuroradiologues experts étaient disponibles et constituaient la « vérité terrain » à laquelle les résultats automatiques étaient comparés. Si les performances restaient inférieures à celles obtenues par le consensus de neuroradiologues, elles en étaient très proches (70 à 80 % d'agrément). La principale difficulté pour les solutions automatiques reste la détection des lésions de petite taille. Tous

les algorithmes étaient transférés et exécutés sur la plateforme, afin de les comparer dans les mêmes conditions. Par ailleurs, les données initiales et les résultats étaient stockés et restent accessibles sur demande sur la plateforme.

Même si actuellement l'imagerie préclinique n'est pas réalisée sur de grandes cohortes, des besoins similaires aux études chez l'Homme se font jour pour assurer la reproductibilité et la robustesse des études statistiques. Ainsi, le concept de « **FAIR** data » (**F**acile à trouver, **A**ccessible, **I**nteropérable, et **R**éutilisable), devient central pour toutes les études en science de la vie. L'ontologie de domaine initialement axée sur l'humain, a donc été étendue pour tenir compte des besoins des études précliniques (concept d'espèce, de modèle animal, anesthésique, ...).

Une nouvelle phase vient de démarrer (2020) qui vise à rendre cette architecture opérationnelle et à évaluer les conditions pour qu'elle remplisse les services attendus par la communauté des chercheurs en imagerie *in vivo*. Une collaboration avec trois entreprises sélectionnées suite à un appel d'offre a été établie pour 24 mois. Il s'agit de mettre en œuvre notre infrastructure de gestion et de traitement de données d'imagerie centrée sur l'utilisation respectivement de Shanoir et de VIP. Les objectifs sont de fournir des services de qualité industrielle : en termes de gestion des comptes utilisateurs, de la maintenance, du support utilisateur ; avec un haut niveau de sécurité et d'hébergement de données de santé ; d'élargir la communauté des utilisateurs ; et enfin, de développer un modèle économique adapté aux ressources financières disponibles dans les centres de recherche publique, pour en assurer la maintenance et l'évolution. Notre portail Internet permettra la recherche et l'accès à des données et outils de traitement, facilitant leur visibilité et leur partage dans le respect des contraintes juridiques et éthiques attachées à la protection des données du vivant (conformité avec le règlement général sur la protection des données RGPD, anonymisation, sécurisation des accès, hébergement de données de santé, etc.). Un nouveau défi pour la détection des lésions de sclérose en plaques aura lieu en 2021 sur la nouvelle version de la plateforme IAM (Information Analysis and Management) pour mesurer, cinq ans après le premier, les avancées des performances, notamment avec l'apparition des algorithmes d'apprentissage automatique.

## **Conclusion**

L'amélioration de la qualité des études publiées en science de la vie passe par la mise à disposition des données et des chaînes de traitement utilisées. Le partage des données permettra d'augmenter la taille des échantillons et donc la robustesse des analyses statistiques, et de revisiter leur analyse par de nouveaux outils de traitement. L'accessibilité des chaînes de traitement facilitera leur diffusion et la reproduction des résultats publiés. La plateforme IAM facilite la visibilité des données d'imagerie *in*



*in vivo* et permet l'exécution des outils de traitement sur une plateforme dédiée. Des travaux sont en cours pour rendre cette plateforme orientée imagerie *in vivo* interopérable avec d'autres bases de données en santé regroupant des données génétiques, cliniques et d'imagerie *in vitro*. Ceci afin d'explorer différentes dimensions du vivant et d'étendre nos connaissances physio-pathologiques.

L'introduction récente pour le traitement des images en biologie-santé des techniques d'apprentissage automatique, grandes consommatrices de données et de moyens de calcul, renforce l'intérêt pour une telle plateforme. Cette adéquation avec l'engouement actuel pour le « big data » et le « machine learning » souligne à la fois la pertinence de travaux ayant démarrés il y a presque vingt ans et la nécessité d'être persévérant pour concrétiser les intuitions initiales et renforcer progressivement les solutions et les choix technologiques déployés.

## Références

M. Bouzeghoub et R. Mosseri (dir) Les Big Data à découvert, CNRS Editions, 2017.

## Remerciements

France Life Imaging bénéficie d'un soutien financier "Investissement d'Avenir" de l'Agence Nationale pour la Recherche (ANR-11-INBS-0006).

## Glossaire

**Algorithme d'apprentissage (*machine learning*).** Différentes techniques statistiques permettent à un algorithme d'adapter progressivement et automatiquement ses paramètres pour améliorer ses performances, généralement de classification. L'apprentissage peut être supervisé, c'est-à-dire que des données exemples sont annotées par un humain ; ou non-supervisé, sans indication précise sur la classe des données. L'apprentissage semi-supervisé est intermédiaire. La puissance des ordinateurs actuels et la mise à disposition de grandes bases de données ont fait récemment exploser leur utilisation, notamment pour la classification d'images médicales.

**Big data.** Terme générique pour désigner l'utilisation de grandes quantités de données.

**Défi.** Une compétition entre fournisseurs de solutions, dans notre cas pour traiter des images.

**Étude multicentrique.** Principalement réalisée en recherche clinique, il s'agit de mettre à contribution différents centres avec un protocole de recueil harmonisé, dans le but d'obtenir assez de données.

**IRM.** Imagerie par Résonance Magnétique. Technique d'imagerie non invasive (sans ajout de produit de contraste) qui repose sur la soumission d'un corps biologique à un champ magnétique statique important. Elle permet d'avoir accès à des informations anatomiques, fonctionnelles ou vasculaires.

Les propriétés magnétiques étant différentes selon la composition des tissus, cela permet d'avoir un contraste dans l'image variable, mettant en évidence l'anatomie, par exemple pour le cerveau, la matière blanche, la matière grise et des zones pathologiques. L'imagerie cérébrale fonctionnelle, qui repose sur les propriétés magnétiques du sang, permet de détecter des zones activées en fonction de stimuli variés. Cela permet d'approfondir nos connaissances sur le fonctionnement normal et pathologique du cerveau. L'IRM de perfusion permet quant à elle d'explorer la vascularisation cérébrale et l'IRM de diffusion les fibres de matière blanche qui connectent les différentes zones cérébrales.

**Mapping.** Mettre en correspondance des concepts présents dans le modèle de référence (comme l'image structurale) et ceux équivalents présents dans un modèle local (comme l'image anatomique).

**Ontologie.** Modèle conceptuel de données qui relie différents concepts importants d'un domaine avec différentes relations (comme « est-un », « composé-de », etc.). Ce modèle représentable dans un langage informatique permet ensuite de faire des raisonnements automatisables.

**Ontologue.** Expert en construction d'ontologies, généralement un informaticien spécialisé en représentation de connaissances.

**Pipeline.** Chaîne de traitement exécutable composée de différents algorithmes pour traiter les images, comme diminuer le bruit dans les images, la taille de l'image ou réaligner des images entre elles.

**RGPD.** Règlement général sur la protection des données qui définit à l'échelle européenne le cadre juridique pour l'utilisation des données personnelles dans le cadre de l'évolution des technologies de l'information.

**Score de comportement.** En neuropsychologie, différentes batteries de tests ont été développées, conduisant à un score pour explorer différentes fonctions cognitives et mémorielles. Ces scores sont utilisés pour raffiner un diagnostic clinique.

**Taxonomie.** Organisation hiérarchique de concepts dans un but de classification. Utilisée dans la construction d'ontologies.

**Tomographie par émission de positrons (TEP).** Technique d'imagerie permettant d'accéder au métabolisme des organes. C'est une technique invasive qui nécessite l'injection d'un traceur. Dans le cas du cerveau et de l'étude des troubles cognitifs, un traceur spécifique permet d'étudier le métabolisme glucidique pour affiner un diagnostic.

## **Affiliation**

Michel Dojat. Neuroimagerie et Neuroinformatique, Inserm – Directeur de recherche, Grenoble Institut des Neurosciences, Grenoble. [michel.dojat@inserm.fr](mailto:michel.dojat@inserm.fr)

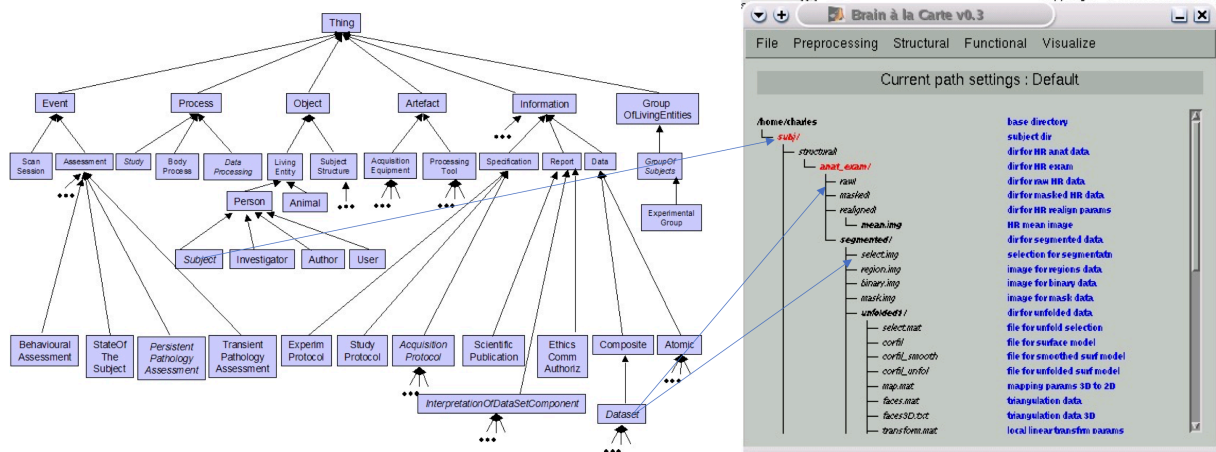


Figure 1. Structuration des données. Gauche : Extrait de la hiérarchie de concepts du modèle public. Droite : Organisation locale des données d'imagerie. Les liens d'équivalence qui doivent être établis entre les concepts référents et les termes locaux sont indiqués par les flèches bleues. Seules les relations de type « est-un » sont indiqués par les flèches noires.

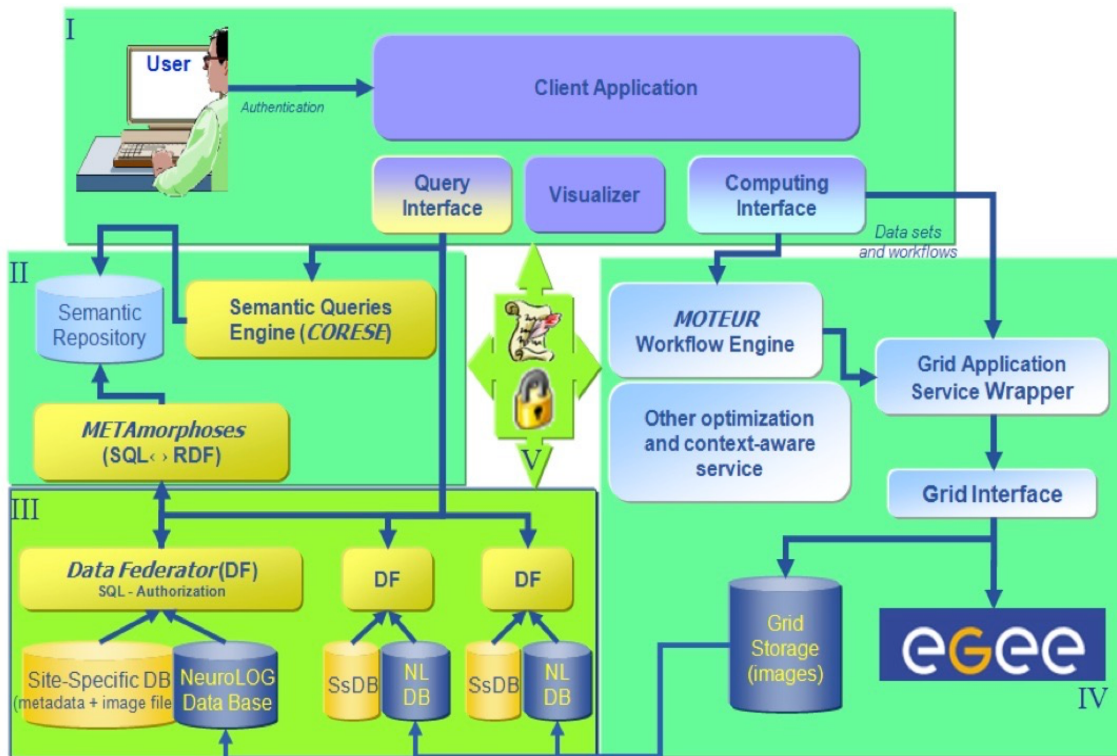


Figure 2. Schéma de l'architecture du système NeuroLOG. I. L'interface utilisateur pour les requêtes, la visualisation, la composition et l'exécution de chaînes de traitement. II. Module sémantique pour le traitement des requêtes vers le modèle de données fédéré public et leur traduction en langage de base de données (SQL). III. Interface avec les bases de données distribuées et de modèles locaux hétérogènes de données. IV. Module de composition de chaîne de traitement, lien vers les données à traitées et pour son exécution sur des plateformes spécifiques (comme EGEE).

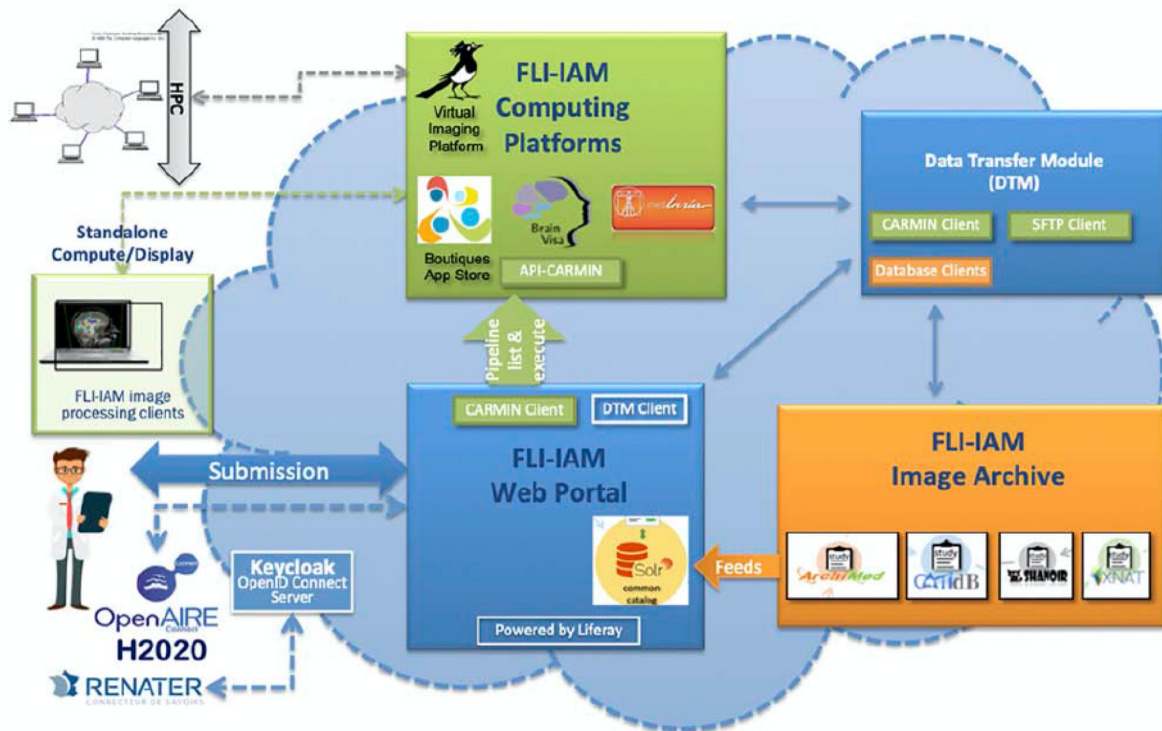


Figure 3. Architecture FLI-IAM. Trois blocs sont au cœur de l'architecture : le portail Internet (en bleu), les bases de données distribuées (en orange), et la plateforme d'exécution des pipelines (en vert). Le module DTM assure les échanges de données entre les blocs. L'API Carmin met en relation données et algorithmes. L'utilisateur soumet une requête *via* le portail (sélection d'images respectant certains critères, choix d'un outil de traitement d'images). Son identification est contrôlée (*via* keycloak). Les traitements d'images peuvent être exécutés localement ou sur des environnements spécifiques (HPC : high performance computing). Le système fait partie d'un projet européen (OpenAire) pour favoriser la science ouverte. D'après Kain et al Front in Neuro 2020.