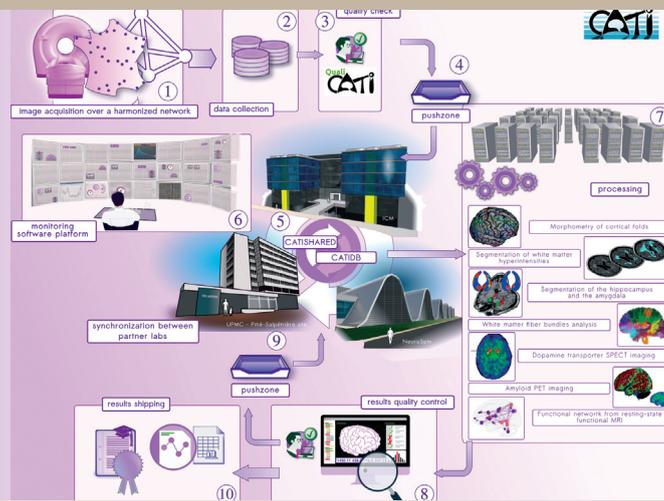


Flux de données au sein du CATI

Après réception par service web, les images sont contrôlées puis envoyées aux équipes d'analyse réparties sur trois sites mais dont la production est synchronisée quotidiennement.



© G. OPERTO

LE CATI HARMONISE LA NEURO-IMAGERIE

Durant la dernière décennie, les problématiques impliquant une collecte multicentrique de données de neuro-imagerie se sont multipliées, qu'il s'agisse de recherche clinique, d'essais thérapeutiques ou de fouille massive de données dans une optique d'intelligence artificielle. Du fait de l'hétérogénéité du parc de scanners, ce besoin achoppe souvent sur l'effort préalable d'harmonisation des acquisitions nécessaire à une exploitation optimale de tels jeux de données. En outre, la complexité logistique de la collecte et de l'analyse de données multicentriques implique généralement des coûts prohibitifs hors de portée de la plupart des études, comme en atteste l'expérience du projet nord-américain précurseur ADNI^{*1}, dans lequel plus de cent millions de dollars ont été investis pour suivre un millier de sujets nord-américains dans le contexte de la maladie d'Alzheimer.

Le problème de l'harmonisation est résolu simplement par des études épidémiologiques sur le vieillissement grâce aux critères d'inclusion très ouverts. Il suffit d'installer des machines identiques dédiées dans quelques grands bassins de population, comme le fait le programme UKbiobank, qui compte 100 000 sujets inclus en IRM (1). Cette approche n'est malheureusement pas adaptée quand des critères d'inclusion plus stricts nécessitent l'utilisation de nombreux imageurs préexistants. On peut néanmoins penser qu'en accumulant des millions d'images, la magie du *big data* opérera pour s'affranchir des effets liés aux scanners. C'est le pari du Human Brain Project, un projet européen phare qui vise entre autres à agréger les images acquises par le passé dans une centaine de grands hôpitaux (2).

Grâce au plan Alzheimer 2008-2012, une troisième voie a été

expérimentée en France. Elle a consisté à profiter du lancement de Memento, une très grande cohorte dédiée à l'étude de l'histoire naturelle de la maladie d'Alzheimer pour construire une plateforme susceptible de répondre aux besoins d'études variées (3). La plateforme CATI^{*2} est ainsi née en 2011 d'un groupe de laboratoires de recherche en neuro-imagerie associant NeuroSpin (CEA), des unités mixtes de recherche Inserm-UPMC-CNRS-Inria, l'Assistance publique-Hôpitaux de Paris et l'Institut du Cerveau et de la Moelle épinière (ICM) à l'hôpital parisien de la Pitié Salpêtrière (4). Elle s'est appuyée sur des experts disséminés partout en France pour construire itérativement ce que les médecins appelleraient un « très grand instrument » dédié à l'imagerie des cohortes, c'est-à-dire un réseau d'une centaine de scanners dont les paramètres d'acquisitions (séquences IRM, paramètres de reconstruction TEP...) ont été réglés de manière à minimiser les effets de site. Ce réseau est constitué, pour les deux tiers, d'IRM et pour un tiers, d'imageurs de médecine nucléaire (TEP ou SPECT^{*3}). Il est aujourd'hui surveillé en permanence pour réagir aux mises à jour matérielles ou logicielles. Les images sont collectées par un service *web* sécurisé et contrôlées par des attachés de recherche clinique en relation avec le réseau d'acquisition.

Booster technologique

Le CATI cherche également à jouer un rôle de concentrateur de technologies d'analyse accélérant la mise à disposition des avancées réalisées en France ou ailleurs. Il valorise ainsi de nombreux logiciels initialement conçus au sein de laboratoires académiques français et mène ses propres programmes de R&D pour minimiser

autant que possible les biais liés à la multiplicité des types de scanners et améliorer la robustesse des algorithmes d'analyse. Il développe également des approches originales permettant d'optimiser voire d'automatiser le contrôle qualité des analyses. Chaque étude choisit au préalable dans un catalogue les biomarqueurs adaptés à ses objectifs. Les images sont ensuite analysées avec un niveau de contrôle qualité et de productivité quasi-industriel. Le CATI cherche ainsi à jouer, pour la recherche clinique académique, un rôle analogue à celui joué par les sociétés de recherche contractuelles dans les essais.

En rendant l'imagerie multicentrique financièrement abordable grâce à des gains d'échelle, le CATI a accéléré l'émergence de nouvelles études cliniques. Aujourd'hui, plus d'une trentaine d'études font ainsi appel à ses services, sur un large spectre de pathologies : maladie d'Alzheimer, démences à corps de Lewy, démences fronto-temporales, maladies de Parkinson et de Huntington, sclérose latérale amyotrophique, troubles bipolaires... Le CATI a déjà analysé les images issues de plus de 10 000 sujets, générant une base de données nationale harmonisée multi-pathologies sans équivalent dans le monde. Cette base sera peu à peu mise à disposition de la communauté du « *machine learning* » (intelligence artificielle) pour nourrir la recherche de biomarqueurs distribués hors de portée des approches statistiques classiques. Il serait en outre aujourd'hui envisageable de capitaliser sur les connaissances accumulées pour proposer des standards d'acquisition utilisables sur l'ensemble du parc d'imageurs français. Une telle harmonisation créerait le terreau nécessaire pour que des abaques normatifs issus des études de recherche soient transférés rapidement à la clinique.

NEUROIMAGERIE

Les auteurs

Jean-François Mangin¹, Marie-Odile Habert^{2,3}, Yann Cointepas¹ et Marie Chupin⁴

¹ Neurospin, Institut Frédéric Joliot, CEA, Université Paris-Saclay, Gif-sur-Yvette

² Service de Médecine Nucléaire, Hôpital Pitié-Salpêtrière, AP-HP, Paris

³ Laboratoire d'Imagerie Biomédicale, Inserm U1146, CNRS UMR7371, Sorbonne Universités, Université Paris 6 Pierre et Marie Curie, Paris

⁴ Sorbonne Universités, Université Paris 6 Pierre et Marie Curie, Inserm, CNRS, Institut du cerveau et la moelle - Hôpital Pitié-Salpêtrière, Paris

- (1) www.ukbiobank.ac.uk
- (2) www.humanbrainproject.eu
- (3) www.memento-cohort.org
- (4) cati-neuroimaging.com

*1 Alzheimer's Disease Neuroimaging Initiative

*2 Centre d'Acquisition et de Traitement d'Images pour la maladie d'Alzheimer

*3 Single photon emission computed tomography : tomographie monophotonique



FLI-IAM : partager les données et les outils de traitement d'imagerie *in vivo*

L'accès distribué aux données et outils de traitement est aujourd'hui indispensable pour une exploitation optimale des informations contenues dans les images acquises *in vivo* dans différents centres, chez l'homme et l'animal. Une architecture informatique permet aujourd'hui la fédération structurée de données hétérogènes, leur fusion, l'exécution de chaînes de traitement et le partage à grande échelle des informations stockées.

© A. NOOR/BSIP

L'époque du *big data* a déjà débuté dans le monde de l'imagerie *in vivo*, où les solutions matérielles et logicielles doivent évoluer pour exploiter efficacement les quantités massives de données disponibles pour les chercheurs, les industriels ou les cliniciens, à l'échelle nationale ou internationale. L'exploitation de ces grandes sources de données est entravée par l'absence ou la fragmentation des infrastructures numériques permettant de les stocker, de les gérer et de les traiter. Par exemple, un clinicien effectuant une étude multicentrique a besoin de solutions pour harmoniser les protocoles d'acquisition d'images, collecter les données sous un seul système d'information, en contrôler la qualité, effectuer l'anonymisation, prétraiter ces données et effectuer des analyses informatisées pour ensuite les partager et diffuser les résultats.

Par ailleurs, une *start-up* qui veut exploiter la découverte d'un nouveau biomarqueur d'imagerie a deux solutions pour évaluer son

innovation à grande échelle : elle peut passer des mois, voire des années, à lever des fonds pour acquérir des données, négocier leurs accès et monter les études multicentriques nécessaires pour valider son concept, ou bien mettre en œuvre seule son processus d'évaluation, en utilisant un ensemble limité de données disponibles dans quelques banques de données ouvertes, et valider sa solution avec des centres cliniques partenaires. Ces exemples pourraient être complétés par de nombreux cas de chercheurs qui luttent pour gérer et traiter leurs données efficacement sur plusieurs plateformes incompatibles.

Dans l'ensemble, la communauté de l'imagerie *in vivo* reconnaît aujourd'hui que le partage des données et des outils de traitement sur ces données d'imagerie est crucial pour la recherche translationnelle, ce que reflètent également les priorités actuelles des organismes de financement de la recherche (1,2). L'objectif du projet FLI-IAM*1 est de fournir une

solution intégrée de gestion des données et des ressources de traitement associées afin de répondre à ce défi. Le déploiement de cette infrastructure nationale ouvrira de nouveaux horizons aux utilisateurs finaux pour mieux tirer parti des nouvelles technologies du numérique dans le domaine de l'imagerie du vivant.

Des plateformes se mettent ainsi en place sous la houlette de différents centres d'excellence scientifique – LONI*2 (3), C-BRAIN*3 (4), COINS*4 (5)... afin de stocker, partager et réutiliser les données biologiques et les métadonnées associées. Pour fusionner et interpréter ces masses de données, des outils de traitement doivent être mis en place, combinés en pipelines et comparés de façon à définir les meilleurs d'entre eux face à un objectif donné. Ces outils doivent être exécutés sur des environnements informatiques adaptés, assurant sécurisation des accès et temps d'exécution minimal. Ce que permettent les avancées de la technologie *web*.

Les auteurs

Michel Dojat^{1,2}, Michael Kain³⁻⁵ et Christian Barillot³⁻⁵

¹ Inserm U1216, Institut des Neurosciences Grenoble

² Université Grenoble Alpes, Grenoble Institut des Neurosciences, Grenoble

³ Inria, Équipe Vision, Action et Gestion d'Informations en Santé (VisAGes), Rennes

⁴ CNRS, UMR 6074 Institut de recherche en informatique et systèmes aléatoires, Université Rennes I, Équipe VisAGes, Rennes

⁵ Équipe VisAGes Inserm U746, Rennes

*1 FLI : France Life Imaging ; IAM : Information Analysis and Management (Analyse et gestion de l'information)

*2 loni.usc.edu

*3 cbrain.mcgill.ca

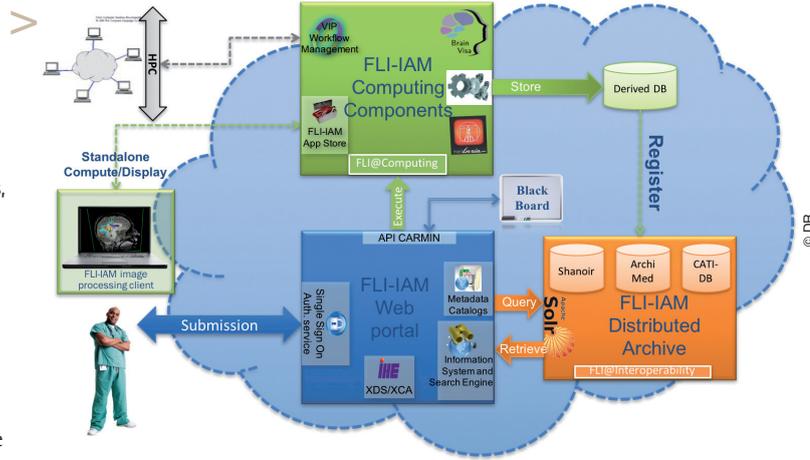
*4 coins.mrm.org

*5 Alzheimer's Disease Neuroimaging Initiative : initiative nord-américaine démarrée en 2004 dont l'objectif est d'améliorer la compréhension de la maladie d'Alzheimer et le développement de traitements en facilitant l'accès aux données longitudinales biologiques, d'imagerie, génétiques et cliniques de patients recueillies selon des protocoles stricts. www.adni-info.org

*6 Medical Image Computing and Computer Assisted Interventions Conference, www.miccai2016.org

FLI-IAM en image

Au travers d'un portail d'accès web sécurisé, un utilisateur, clinicien ou chercheur, peut accéder à la fois à des bases de données distribuées dans différents centres (Shanoir à Rennes, ArchiMed à Nancy, Cati-DB à Orsay) pour récupérer ou stocker des données d'imagerie. Il accède aussi à un catalogue (FLI-IAM Appstore) d'outils de traitement de référence. L'exécution de ces outils sur des ressources HPC (*high performance computing*) génère des données dérivées de qualité, stockées localement ou dans les bases de données accessibles. L'infrastructure peut fédérer de nouvelles bases de données distribuées et interopérer avec des infrastructures internationales existantes.



L'INFRASTRUCTURE FLI-IAM

Dans le cadre du programme d'Investissements d'avenir *Infrastructure en biologie et santé*, FLI a pour ambition de coordonner, à l'échelle nationale, les activités de recherche et les compétences en imagerie *in vivo* chez l'homme et l'animal, et de fournir un accès à un éventail complet de technologies d'imagerie – riche de 150 systèmes – et de services intégrés et harmonisés.

Au sein de FLI, le projet IAM a pour objectif la mise en place d'une infrastructure informatique, logicielle et matérielle, pour la gestion et le traitement dématérialisé des données d'imagerie *in vivo*, chez l'homme et le petit animal, en provenance d'acteurs très divers (centres de recherche, hôpitaux, recherche clinique publique et privée...).

FLI-IAM propose une infrastructure versatile, développée à partir de composants logiciels fournis par les partenaires académiques du projet et permettant la connexion de composants, allant de la gestion des données et métadonnées d'imagerie au traitement des images produites, à travers des outils mis à disposition sur le poste de travail de l'utilisateur ou à travers l'usage de ressources de capacité massive de calcul (grilles de calcul, *clusters*...) (figure ci-dessus).

FLI-IAM supporte trois types de familles de scénario d'usage :

Scénario 1, de type « Software-as-a-service »

Ce scénario issu du monde de l'informatique « en nuage » (*cloud computing*) fournit des services pour des usages de type recherche clinique et recherche préclinique multicentrique. L'infrastructure

supporte tous les niveaux technologiques pour la fourniture de ces services, du réseau de communication aux outils d'exploitation pour l'utilisateur, en passant par les infrastructures de calcul et de stockage et la gestion/maintenance des progiciels fournissant le service.

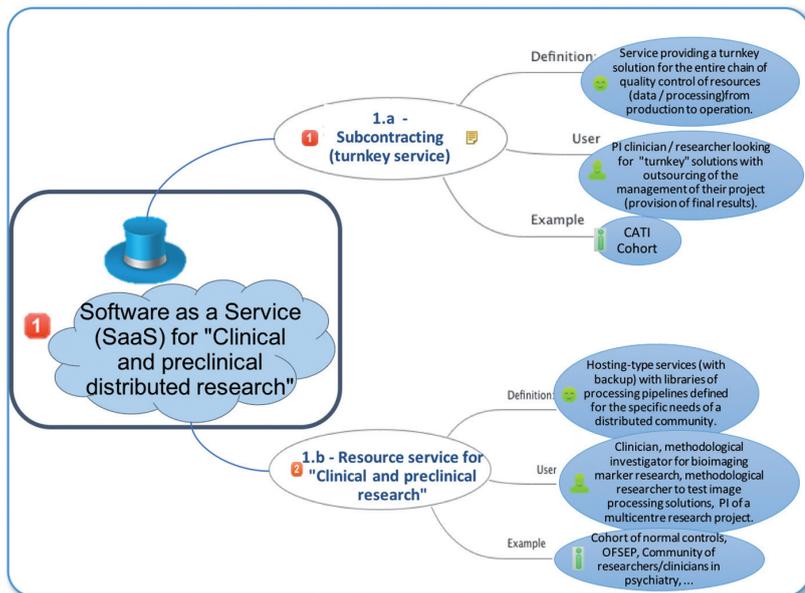
Deux niveaux de service sont disponibles : « clé en main » et « à façon » (figure ci-dessous). Dans le premier cas, l'utilisateur n'a pas d'accès direct aux outils de l'infrastructure mais uniquement aux résultats du service demandé, allant de l'harmonisation des protocoles d'imagerie à la fourniture d'un rapport d'analyse des données acquises, en passant par le recueil et le contrôle qualité. Le service « à façon » offre la mise à disposition d'espace de stockage, d'accès aux données et à un catalogue d'outils de traitement. Le premier niveau de service est orienté vers la prise en charge partielle ou totale des étapes d'un projet multicentrique d'imagerie *in vivo*, le second, vers l'assistance à sa réalisation pratique via l'accès à l'infrastructure, dont les membres restent alors les acteurs centraux.

Scénario 2, de type « Platform-as-a-service »

Ce scénario fournit une solution « *Open Resources* » afin de mettre à la disposition d'une large communauté des données ou des solutions de traitement d'images. L'infrastructure met à disposition les couches basses nécessaires à la fourniture de services, du réseau de communication aux progiciels hébergeant les services fournis par l'utilisateur, en passant par les infrastructures de calcul et de stockage et la gestion/maintenance des progiciels. Dans ce scénario, l'utilisateur – y compris l'utilisateur fournisseur de service – pourvoit à la solution métier/service final et utilise l'infrastructure pour héberger son service. L'utilisateur peut, par exemple, être une *start-up* ayant développé une solution innovante et désirant l'exposer à des fins de tests et de validation à une large communauté sur des données ouvertes et mise à disposition par la communauté de l'infrastructure. On peut envisager la mise à disposition de données produites

Scénario 1, de type « Service as-a-software »

Deux types de services sont disponibles : « clé en main », où différentes étapes du projet sont entièrement gérées par l'infrastructure, et « à façon », où l'utilisateur recherche une assistance pour la mise à œuvre de son projet, dont il reste l'acteur central.



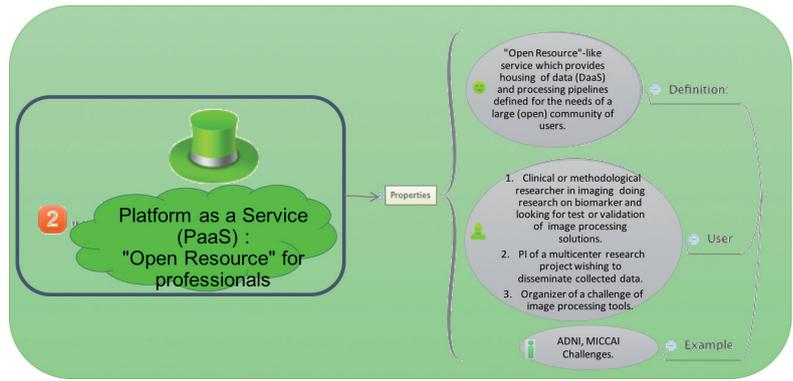
dans le scénario 1 pour un usage ouvert, comme celui de la cohorte ADN15 (figure ci-contre).

Scénario 3, de type « Storage as a service »

Ce scénario fournit une solution « cloud » de stockage et de partage de copies de ressources locales en ligne sans structuration imposée et dans un environnement sécurisé pour les données de santé. Il permet à des usagers de partager de manière libre et respectant la réglementation (confidentialité) des ressources numériques à travers un hébergeur virtuel (figure ci-dessous).

FLI-IAM EN PRATIQUE

Dans le cadre de service en lien avec le scénario 2, l'infrastructure a été utilisée récemment pour organiser avec succès deux compétitions de traitement d'images médicales pour la segmentation de la charge lésionnelle dans des IRM de patients atteints de sclérose en plaques (compétition MSSEG) ou des lésions tumorales dans des images en TEP (compétition PETSEG), lors de la conférence internationale MICCAI⁶, qui s'est tenue en octobre 2016 à Munich. Dans les deux cas, le scénario 2 a été utilisé pour partager avec les différents compétiteurs – n=14 pour MSSEG, n=7 pour PETSEG – des données d'apprentissage et la « vérité terrain » – la délimitation manuelle des lésions par différents experts – pour que chaque équipe puisse vérifier la pertinence de sa solution. Dans un deuxième temps, les différentes solutions concurrentes ont été portées sur l'infrastructure puis exécutées face à des nouveaux jeux de données d'imagerie. Les résultats obtenus, stockés sur l'infrastructure, ont permis le calcul de scores quantitatifs puis l'établissement d'un classement des méthodes. Tous les résultats étaient consultables par les compétiteurs ainsi que les temps de calcul déterminés après exécution sur la même infrastructure sans aucune intervention des compétiteurs. Les données d'apprentissage présentes sur l'infrastructure sont aujourd'hui à la disposition de la communauté scientifique.



Scénario 2, de type Platform-as-a-service

L'infrastructure est mise à disposition pour héberger des services, par exemple un composant logiciel à tester ou des données à partager. Autre exemple d'utilisation : la réalisation de compétitions pour valider et comparer des solutions face à un objectif commun.

UNE INFRASTRUCTURE PARTAGÉE

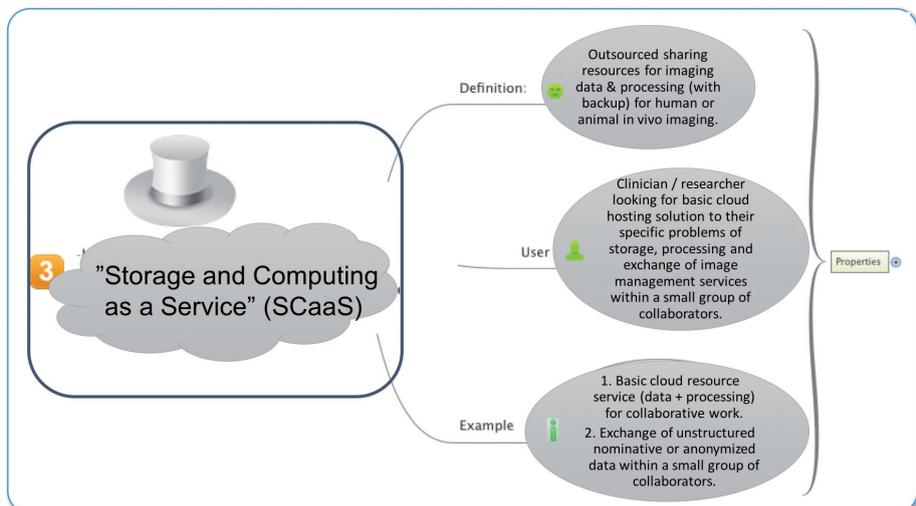
Par ailleurs, l'infrastructure FLI-IAM est utilisée pour partager des données issues d'un petit animal (souris) entre cinq centres – Grenoble Institut des Neurosciences (GIN), le Centre de Résonance Magnétique Biologique et Médicale de Marseille (CRMBM), le Molecular Imaging Research Center (MIRcen) et NeuroSpin, à Saclay, la Fédération de Recherche en Imagerie Multimodalité (FRIM), à Paris, et le Laboratoire des sciences de l'Ingénieur, de l'Informatique et de l'Imagerie (ICube) de Strasbourg – afin de comparer la qualité des données entre imageurs à valeur identique ou

différente – de 4,7 à 11,4 Tesla – de champ magnétique et les résultats de chaînes de traitement pour l'obtention de cartes paramétriques de relaxation.

L'infrastructure FLI-IAM représente ainsi une avancée technologique importante pour fédérer de façon structurée les données et métadonnées d'imagerie *in vivo* acquises sur différents sites, fusionner des données hétérogènes (structurales, fonctionnelles, comportementales et génétiques), faciliter le déploiement et l'exécution de chaînes de traitement de référence et permettre le développement de projets à large échelle chez l'homme et l'animal pour dépasser les limitations des études actuelles. ■

Scénario 3, de type Storage-as-a-service

L'infrastructure permet l'hébergement et le partage de données dans un environnement sécurisé respectant la réglementation sur les données de santé.



- (1) Ioannidis JP (2014) *PLoS Med* 11(10), e1001747
- (2) Poline JB et al. (2012) *Front Neuroinform* 6, 9
- (3) Dinov I et al. (2010) *PLoS ONE* 5(9), e13070
- (4) Das S et al. (2011) *Front Neuroinform* 5, 37
- (5) Scott A et al. (2011) *Front Neuroinform* 5, 33