

plateforme imageries du vivant
PARIS DESCARTES

7 décembre 2017

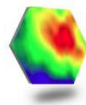
Gestion des données scientifiques en imagerie *in vivo*

Colloque de la plateforme d'imageries du vivant de
Paris Descartes



Centre de Recherche Cardiovasculaire de Paris

avec le soutien  FEALINX



Les acteurs de l'imagerie préclinique sont impliqués depuis plusieurs années dans des démarches de **structuration des résultats** de leur activité. Les enjeux pour les centres de production concernent la **mise à disposition** des données et leur **sécurisation**, la mise en place de **bases de données massives** impliquant des données à la fois variables et évolutives.

Face à la complexité et la multiplicité des sources de données, la démarche organisationnelle **DRIVE** (Déploiement du Réseau d'Imageries du Vivant), fondée sur les outils industriels du Product Life cycle Management (PLM) a été mise en place en cours dans l'environnement de la plateforme d'imageries du vivant de l'Université Paris Descartes afin d'améliorer les échanges et la structuration à l'échelle locale. Un enjeu majeur concerne la capacité des acteurs d'imagerie à **partager** les **données**, les **outils** matériels et logiciels d'analyse, ainsi que les **ressources** informatiques, pour constituer des dépôts. Cette démarche est en cours actuellement dans le cadre du regroupement des universités (projet interdisciplinaire d'Imageries du Vivant **IDV**), avec l'achat de ressources matérielles, la mise en place de machines virtuelles (Cumulus) et la constitution d'une preuve de concept de dépôt multimodal. A une autre échelle, les plateformes d'imagerie se structurent sous l'égide d'organisations nationales comme FLI (France Life Imaging) ou Ibisa (Infrastructures en Biologie et Santé) en vue d'une amélioration et d'une homogénéisation des pratiques (qualité, traçabilité).

Ces démarches témoignent de l'actualité de la gestion des données scientifiques par les communautés de l'imagerie. La plateforme d'Imageries du Vivant de l'Université Paris Descartes est heureuse de vous accueillir à sa journée de débat sur la Gestion des données scientifiques en imagerie in vivo.

Remerciements

Les organisateurs sont reconnaissants aux intervenants de cette Journée d'avoir accepté de donner leur temps pour partager leurs savoir-faire et leurs expériences.

Ils remercient Alain TEDGUI, le directeur du PARCC et Philippe COUDOL, son secrétaire général, d'avoir mis à la disposition du colloque les infrastructures du Centre de Recherche Cardiovasculaire de Paris. Une reconnaissance toute spéciale et un grand merci pour l'organisation matérielle à Mme Karima ICHEGOUR, qui a parfaitement assuré la logistique de l'organisation du colloque, ainsi qu'à Mmes Catherine TRITSCHER et Muriel DELACROIX pour leur aide et leur soutien constants.

Enfin, ils tiennent à remercier particulièrement le soutien de la société FEALINX, sans laquelle ce colloque n'aurait pu avoir lieu.

Daniel BALVAY, Bertrand TAVITIAN

8h45 ACCUEIL DES PARTICIPANTS

INTRODUCTION

9h15 Bertrand TAVITIAN (PARCC - HEGP). Accueil

I – GESTION ET ARCHIVES DES DONNEES NUMERIQUES

9h30 Francis ANDRE (DIST, CNRS). La gestion des données de recherche : contextes politique et scientifique, européen et international.

10h00 Marie-Christine JACQUEMOT-PERBAL (INIST). OPIDor, des services pour faciliter l'adoption de bonnes pratiques de gestion des données de la recherche.

10h30 PAUSE CAFE

II – OUTILS DE DEPOT ET DE CALCULS DE DONNEES SCIENTIFIQUES

11h00 Mickael KAIN (INRIA - Nœud IAM de FLI). Mise à disposition des ressources de calcul pour les communautés de biologie et d'imagerie : les points d'accès aux grilles de calcul. Présentation du Nœud IAM.

11h30 Dimitri VISVIKIS (FLI). Gestion des données pour une meilleure exploitation des données (modélisation multiparamétrique, combinaison biomarqueurs d'imagerie, génétiques, sémantiques, etc). L'expérience de FLI-WP4.

12h00 Christophe CERIN (IDV, Université Paris XIII). Plateformes de ressources numériques : dépôts de données, VM et accès aux ressources de calculs à l'échelle de la communauté d'Universités. L'expérience USPC.

12h30 PAUSE DEJEUNER

III – EXPERIENCES EN IMAGERIE

14h00 Bastien RANCE (APHP - Université Paris Descartes). L'imagerie clinique, états des lieux et enjeux à venir, l'hôpital et la recherche clinique.

14h30 Marc JOLIOT (CEA-CNRS - Université de Bordeaux). Product Life Management en imagerie cérébrale, retour d'expérience du programme ANR Biomist.

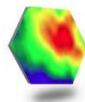
15h00 PAUSE CAFE

15h30 Marianne Allanic (Fealinx). Product Lifecycle Management pour la gestion des données hétérogènes et complexes de la plateforme PIV : le projet DRIVE.

16h00 Yves Frapart pour Philippe GARTEISER (IDV, Inserm). Projet de partage d'images multimodales IDV. Preuve de concept à USPC.

CONCLUSION

16h30 Table ronde animée par Bertrand TAVITIAN et Daniel BALVAY (PARCC - HEGP). Vue d'ensemble, enjeux pour les labos, moyens et actions à mettre en place...



La *gestion des données scientifiques* est un élément moteur de la connaissance scientifique reconnu par de nombreuses instances nationales et internationales (Inserm [1], NIH [2], Horizon H2020 [3], Force 11 [4], ...). Elle est appelée à prendre une importance aussi grande dans la recherche en *imagerie in vivo* que celle qu'elle occupe en imagerie médicale.

Comme d'autres domaines de la connaissance, l'imagerie médicale constitue de grands *dépôts de données* consultables par la communauté scientifique [5] qui deviennent à leur tour de nouveaux matériaux d'études. A ces dépôts de données s'ajoutent de *nouveaux types de ressources* accessibles à la communauté : *bases de connaissances* pédagogiques et scientifiques, *formations* et contenus d'articles en lignes [6], *ressources matérielles* pour le stockage et le calcul intensif par grilles de calcul [7][8] *ressources en logiciels* partagés, et de nombreux outils libres ou payants dédiés aussi bien à l'analyse des données et à la mise en œuvre de *travaux collaboratifs* qu'à la *structuration* des données en vue de leur consultation et de leur archivage [9][10].

L'écosystème scientifique de l'imagerie in vivo est en pleine mutation : pour l'utilisateur un nouveau *substrat pour la recherche*, pour le producteur de *nouveaux formats numériques d'édition scientifique*, dûment identifiés et intégrables comme objets de citation (DOI). L'écosystème en devenir permettra de travailler sur des masses de données bien supérieures aux capacités de production de chaque centre de recherche, réalisant un *changement d'échelle avec optimisation des ressources* (humaines, animales, financières, temporelles...), et limitant la *volatilité des données* post-publication. Dans cet écosystème, le *croisement d'informations* ouvre de nouvelles perspectives de pratique de la recherche.

La mise à disposition des résultats d'études à leurs différents stades d'avancement permet d'augmenter la *reproductibilité des données*, donc leur crédit. Cette *valorisation* implique un *effort éditorial* pour rendre lisibles les données et les méthodes. Que ce soit pour un usage *interne* ou pour un *dépôt public*, cet effort s'ajoute à celui investi lors d'une publication classique dans une revue scientifique. La structuration des données et des pratiques, en vue d'un archivage efficace, rejoint les questions de *transparence*, de bonne gestion et de *qualité*. Elle tend à encourager rigueur scientifique et *diffusion de l'état de l'art*. Accéder aux informations à leur source permettra des *méta-analyses* de meilleure qualité et de plus grande valeur.

L'évolution en cours pose des questions très concrètes : quelles *règles* et quelles *bornes* établir dans les *partages* de ressources ? A *qui appartiennent-elles* ? Qui est *compétent* pour les gérer ? Quelle formation pour les acteurs ? S'agit-il d'un *nouveau cœur* de métier ? Quels sont les *risques* associés à la transparence ? Quelle sera l'acceptabilité de cette évolution par ses acteurs ? A quelle échelle structurer l'activité ? A quel niveau devront se prendre les décisions ? Quel équilibre entre l'investissement et la valeur ajoutée ? Quel sera le *coût associé* à cette nouvelle structuration ?

[1] Inserm Workshop 246 : *Management and Reuse of health data : methodological issues* [2017]

[2] NIH : <https://datascience.nih.gov/index>

[3] H2020 Programme : *Guidelines on FAIR Data Management in Horizon 2020*

[4] Force 11 : <https://www.force11.org/about>

[5] TCGA Data Portal : <https://cancergenome.nih.gov/>

[6] MOOCS : <https://mooc-francophone.com/> ; Coursera

[7] Grilles de calculs : <http://in2p3.fr/>

[8] Grille du CERN <http://home.cern/fr/about/computing/worldwide-lhc-computing-grid>

[9] PACS : https://en.wikipedia.org/wiki/Picture_archiving_and_communication_system

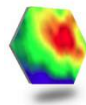
[10] Gestion : https://fr.wikipedia.org/wiki/Management_du_système_d'information

La gestion des données de recherche : contextes politique et scientifique, européen et international

Francis ANDRE

*DIST/CNRS, 3 rue Michel-Ange 79794 Paris Cedex 16
francis.andre@cnrs-dir.fr*

Les données de la recherche sont des éléments essentiels du mouvement international vers la science ouverte. Leur diversité, hétérogénéité et variabilité observées dans les différents environnements disciplinaires amènent à construire des modèles de gestion de données respectueux des pratiques des communautés scientifiques. Mais à une époque où la recherche est de plus en plus massivement numérique, coopérative et interdisciplinaire, la mutualisation des infrastructures de recherche apparaît comme une évidence. Dans ce schéma apparemment contradictoire, la présentation analysera les différents paramètres et mécanismes conduisant à une meilleure gestion des données de la recherche au bénéfice d'une recherche et innovation plus efficaces.



OPIDoR, des services pour faciliter l'adoption des bonnes pratiques de gestion des données de la recherche

Marie-Christine Jacquemot-Perbal

Inist-CNRS, 2 allée du Parc de Brabois, 54519 Vandoeuvre lès Nancy, France.

L'Open Science a pour ambition de rendre la recherche encore plus crédible, fiable, efficace et plus réactive pour répondre aux défis sociétaux. Dans ce but, les données doivent être partagées le plus librement possible, dans des formats qui permettent leur intégration et analyse automatiques quelle que soit leur provenance disciplinaire ou géographique. En jargon informatique, cela se traduit par le terme « d'interopérabilité » des données.

L'acheminement vers cet objectif est guidé par les principes FAIR qui facilitent la découverte, l'accès, l'interopérabilité et la réutilisation des données. Leur mise en œuvre requiert l'engagement des chercheurs mais aussi des services de soutien à la recherche qui s'emploient à lever progressivement les obstacles technologiques et juridiques de la collecte/génération des données jusqu'à leur réutilisation.

OPIDoR : Optimisation du Partage et de l'Interopérabilité des Données de la Recherche est un projet porté par l'Inist-CNRS. Il a pour objectif de mettre à disposition de la communauté de l'Enseignement Supérieur et de la Recherche des outils et des services facilitant la gestion, diffusion et réutilisation des données en conformité avec les principes FAIR. Deux services vous seront notamment présentés : [DMP OPIDoR](#) et [Cat OPIDoR](#).

France Life Imaging (FLI) - Information Analysis and Management (IAM), entry and access point to resources for medical imaging research

Michael Kain

INRIA Rennes, Campus Universitaire de Beaulieu, 35042 Rennes Cedex - France.

IAM (Information Analysis and Management) is the computation resources node of France Life Imaging (FLI). It provides access to multiple imaging databases and computation resources and takes care of the interoperability between databases¹ and processing pipelines (local or cluster-based platforms). The presentation of FLI-IAM will therefore start with a short introduction and an architecture overview (fig. 1). Afterwards the interoperability between databases (ArchiMed, CATIDB, Shanoir, XNAT) and processing pipelines will be detailed on some examples: CARMIN² (common web API for remote pipeline execution), OAK (open app catalog), VIP³ and Boutiques⁴. The presentation finishes with a short outlook on phase 2 of FLI-IAM.

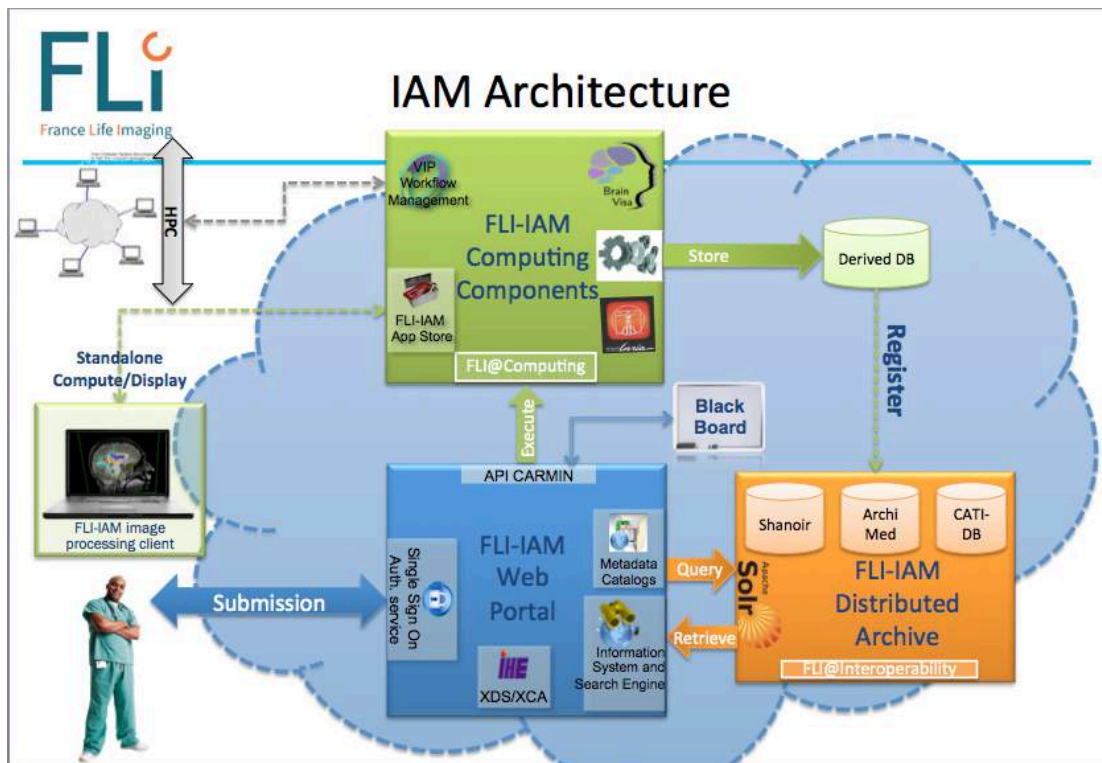
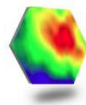


Figure 1. Architecture overview

References

1. Dojat M, Kennedy DN and Niessen W (2017) Editorial: MAPPING: Management and Processing of Images for Population Imaging. *Front. ICT* 4:18. doi: 10.3389/fict.2017.00018
2. Glatard T, Cointepas Y, Commowick O, Kain M, Laurent B, Leray F and Barillot C (2015). CARMIN: a common web API for remote pipeline execution. *Front. Neurosci. Conference Abstract: Neuroinformatics 2015*. doi: 10.3389/conf.fnins.2015.91.00053



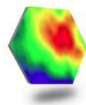
3. Tristan Glatard, Carole Lartizien, Bernard Gibaud, Rafael Ferreira da Silva, Germain Forestier, et al.. A virtual imaging platform for multi-modality medical image simulation.. IEEE Transactions on Medical Imaging, Institute of Electrical and Electronics Engineers, 2013, 32 (1), pp.110-8. doi: <10.1109/TMI.2012.2220154>
4. Glatard T, Da Silva RF, Boujelben N, ADALAT R, Beck N, Rioux P, Rousseau M, Deelman E and Evans AC (2015). Boutiques: an application-sharing system based on Linux containers. Front. Neurosci. Conference Abstract: Neuroinformatics 2015. doi: 10.3389/conf.fnins.2015.91.00012

Multi-parametric multi-modality imaging for prognostic and predictive modeling in oncology

Dimitri VISVIKIS

Director of Research, National Institute of Health and Medical Sciences, INSERM UMR1101, LaTIM, CHU Morvan, Brest (dimitris@univ-brest.fr)

Several quantitative features can be extracted from multimodality images, such as tumor volume, tumor shape characterization, intra-tumor intensity heterogeneity. Some of these features, calculated from baseline multimodality images, have shown substantial potential in improving prognostic and predictive value relative to other patient relative contextual information or disease grade. It has been hypothesized that these features highlight underlying tumor patho-physiological processes at smaller scales. On the other hand, in the field of molecular biology there is increasing interest in the use of different disease markers which can provide complementary information concerning alterations of signaling pathways. The integration of all this heterogeneous information represents a major challenge in the field of multi-parametric modeling with the potential to further improve the selection of patient specific treatment regimens leading to an associated improved overall patient outcome. The challenge resides in the development of methodology for the combination of this multi-scale heterogeneous information and its exploitation in clinical practice. This presentation will demonstrate the potential interest of multimodality multi-parametric imaging for oncology applications.



Plateformes de ressources numériques : dépôts de données, VM et accès aux ressources de calculs à l'échelle de la communauté d'Universités : l'expérience USPC.

Christophe Cerin

Programme Imageries du Vivant, Université Paris XIII

Dans cette présentation nous discuterons des évolutions récentes dans les plateformes de type HPC (High Performance Computing) versus plates-formes pour l'éco-système du Bigdata. Nous essayerons de dégager des points de convergence et/ou de résonance et nous illustrerons notre propos par l'intermédiaire de projets en développement. Il s'agit de nous interroger aussi sur les besoins présents et futurs en matière de grands équipements en informatique, nous entendons par là des équipements avec plusieurs centaines de milliers de processors, voire des millions de processeurs. Nous nous interrogerons en termes de performance, d'énergie consommée, de productivité, de modèle de programmation, de portabilité.

L'imagerie clinique, états des lieux et enjeux à venir, l'hôpital et la recherche clinique.

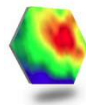
Bastien Rance,

INSERM, AP-HP, France.

t

We live in a world of data. Humans have never accumulated more data in History. In medicine, the growing use of Electronic Health Record, Laboratory Management Software and Picture Archiving and Communication Systems have generated massive amount of facts, largely stored in clinical data warehouses and dedicated storage systems. Paradoxically, these gold mines have largely been unexplored, and fantastic opportunities remain for scientific discovery fueled by evolutions in the field of computer sciences and mathematics.

In our presentation, we will discuss initiatives to integrate clinical imaging, molecular and clinical data, and the rise of new methodologies (e.g. radiomics). We will also describe the hurdles associated with the secondary use of clinical imaging, including the difficulty of associating clinical annotation to clinical images, the issue of interoperability, and consider possible solutions.



Genèse, application et futur de BIOMIST : un système de gestion du « cycle de vie » de la donnée biomédicale

Marc Joliot, CEA

Groupe d'imagerie neurofonctionnelle, IMN, UMR5293, CNRS-CEA-UB, Bordeaux, France

Depuis sa création en 1989 à Orsay, le Groupe d'Imagerie Fonctionnelle (GIN) aujourd'hui installé à Bordeaux a toujours été impliqué dans l'acquisition et l'analyse de données d'imagerie cérébrale sur de grandes cohortes de sujets. Cette activité nous a conduit à développer les outils de gestion d'importantes quantités de données en assurant leur sécurité et leur accessibilité pour une communauté scientifique non spécialiste en technologie de l'information. Fruit de 20 ans d'expérience, le projet BIOMIST a été initié en collaboration avec le service de recherche et d'innovation de l'entreprise Fealinx. Il est apparu dès la genèse de ce projet que les besoins de la communauté de neuroimagerie étaient aussi ceux de l'industrie manufacturière, à savoir : gérer l'ensemble du « cycle de vie » de larges volumes de données hétérogènes, de leur acquisition à leur traitement jusqu'à la production scientifique. Cette présentation abordera les adaptations et développements qui ont été réalisés pour répondre aux exigences de la recherche en bio-imagerie. Deux applications utilisant des données humaines seront présentées pour illustrer les fonctionnalités de BIOMIST. Une présentation des évolutions futures, fruit du retour d'expérience des utilisateurs et de l'évolution du domaine vers la création de bio-marqueurs, conclura cet exposé.

Product Lifecycle Management pour la gestion des données hétérogènes et complexes de la plateforme PIV : le projet DRIVE

Marianne Allanic

R&D, Fealinx, 13 rue de la Loire, 44232 Saint Sébastien-sur-Loire, France

Le projet DRIVE, financé par Sorbonne Paris Cité, a pour objectif de fournir une solution harmonisée pour la gestion des données et de la traçabilité des études biomédicales multimodales de la plateforme PIV de l'université Paris Descartes.

La plateforme SWOMed déployée au sein du PARCC-HEGP s'appuie sur un système de Product Lifecycle Management (PLM) et sur les développements réalisés dans le cadre du projet ANR BIOMIST (Allanic et al., 2017). Grâce au PLM biomédical, quelque soit leur source et leur format, les données des chercheurs sont gérées tout au long du cycle de vie des études, de la planification du protocole expérimental jusqu'à leur valorisation (publication). Les enjeux du projet DRIVE portent sur l'adaptation d'un système conçu pour la gestion d'études cliniques (projet BIOMIST) à la gestion d'études précliniques (plateforme PIV), avec inclusion de nouvelles modalités de données. Dans cette optique l'organisation des données au sein du PLM biomédical a été étendue et une méthode d'import générique de données a été développée. Les premiers résultats du projet DRIVE ont été présentés cette année dans les conférences SMAPP et SIIM (Raboudi et al., 2017). La plateforme est hébergée sur le cloud universitaire mutualisé CUMULUS de Sorbonne Paris Cité, et pourra à terme être mise à disposition de l'ensemble des chercheurs pour faciliter l'échange et la réutilisation de données entre équipes.

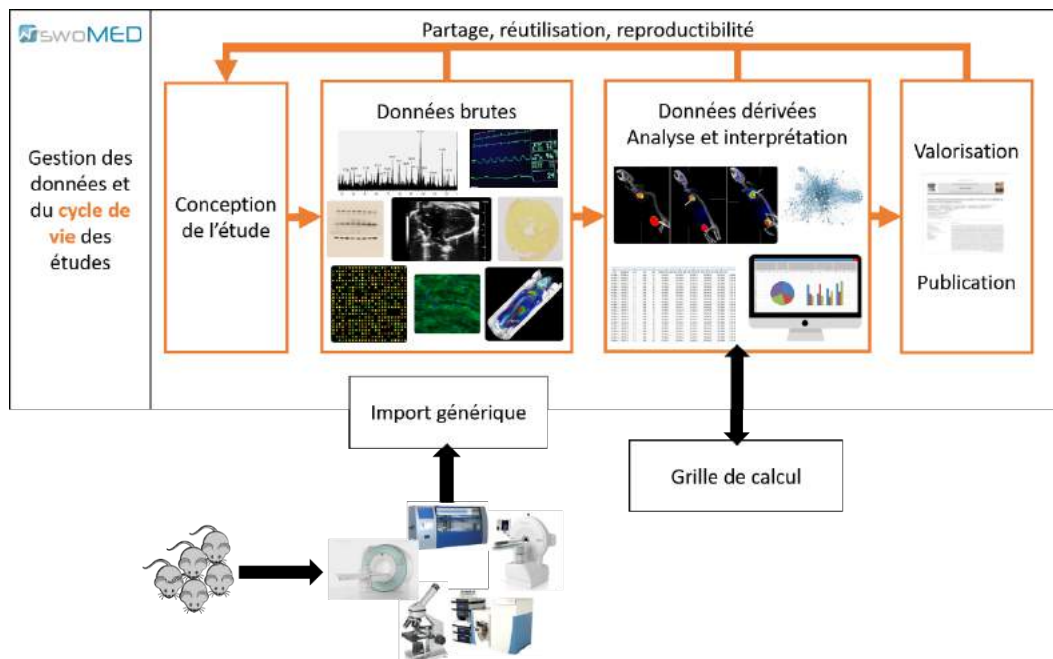
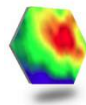


Schéma de la plateforme du projet DRIVE

1. Allanic, M., et al. "BIOMIST: A platform for Biomedical Data lifecycle Management of neuroimaging cohorts." *Frontiers in ICT*, 2017, vol. 3, p. 35.
2. Raboudi A. et al., "Traçabilité de l'intégration de données biomédicales hétérogènes dans le système SWOMed de gestion du cycle de vie des études biomédicales." *Actes du Symposium sur l'Ingénierie de l'Information Médicale (SIIM)*, 23-24 novembre 2017



Le programme interdisciplinaire Imageries Du Vivant de Sorbonne Paris Cité

Philippe Garteiser¹, Danielle Geldwerth-Feniger², Sébastien Li-Thiao-Té³, Yves Frapart⁴, Charles-André Cuenod⁵

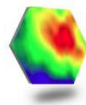
¹Laboratoire des Biomarqueurs en Imagerie, Centre de Recherche de l'Inflammation. Inserm, Paris Bichat, ²Chimie, Structures et Propriétés de Biomatériaux et d'Agents Thérapeutiques (CSPBAT) Université Sorbonne Paris Cité (USPC), Institut Galilée, université Paris 13, CNRS : UMR7244 Institut Galilée, Université Paris 13, 99 avenue Jean-Baptiste Clément, F-93430 Villetaneuse - France, ³Université Paris 13, ⁴Laboratoire de Chimie et Biochimie Pharmacologiques et Toxicologiques (LCBPT) CNRS : UMR8601, Université Paris V - Paris Descartes 45 rue des Saints-Pères 75006 Paris - France, ⁵Paris-Centre de Recherche Cardiovasculaire (PARCC) Hôpital Européen Georges Pompidou [APHP] : U970, Université Paris Descartes - Paris 5 : U970, Institut National de la Santé et de la Recherche Médicale : U970 56 rue Leblanc 75015 PARIS

Imageries du Vivant (IDV), un des neufs programmes interdisciplinaires de Sorbonne Paris Cité, a comme objectif de fédérer les laboratoires et plateformes impliqués dans l'imagerie biologique et médicale au sein de l'Université Sorbonne Paris Cité, en favorisant la communication, les échanges et les collaborations entre ces équipes. En effet, USPC dispose d'un grand nombre d'équipes impliquées dans l'imagerie, dotées d'un plateau technique d'envergure et diversifié incluant des technologies de pointe dans le domaine de l'imagerie du vivant, avec notamment, le premier prototype mondial TEP-Échographie (Parcc-HEGP / Institut Langevin), le premier TEP-IRM haut champ petit animal en région parisienne (P7 Bichat-Beaujon; installation 2017), le premier appareil d'imagerie RPE (Résonance Paramagnétique Électronique) du petit animal et premier appareil de RPE pour humain en France (3 prototypes dans le monde), et l'imagerie à l'échelle nanométrique (Université Paris 13, Bobigny). Cet ensemble totalise plus de 300 chercheurs ou enseignants-chercheurs représentant l'ensemble des domaines impliqués en imagerie du vivant (chimie, physique, biologie, médecine, mathématiques, informatique, sciences sociales). Les appareils et les compétences étant disséminés sur l'ensemble du périmètre de l'USPC, la création de réseaux de communications entre chercheurs pour permettre les échanges intellectuels et la transmission des images entre sites s'avère indispensable pour exploiter et potentialiser cette richesse. Le programme permet également l'élaboration d'une réflexion sur les problèmes éthiques et juridiques soulevés par l'utilisation et le partage des images.

Le programme interdisciplinaire IDV a permis une structuration et animation du réseau IDV, la création d'un environnement informatique de collaboration appelé "Cloud IDV". Ce réseau permet notamment le partage d'espaces de travail au moyen de machines

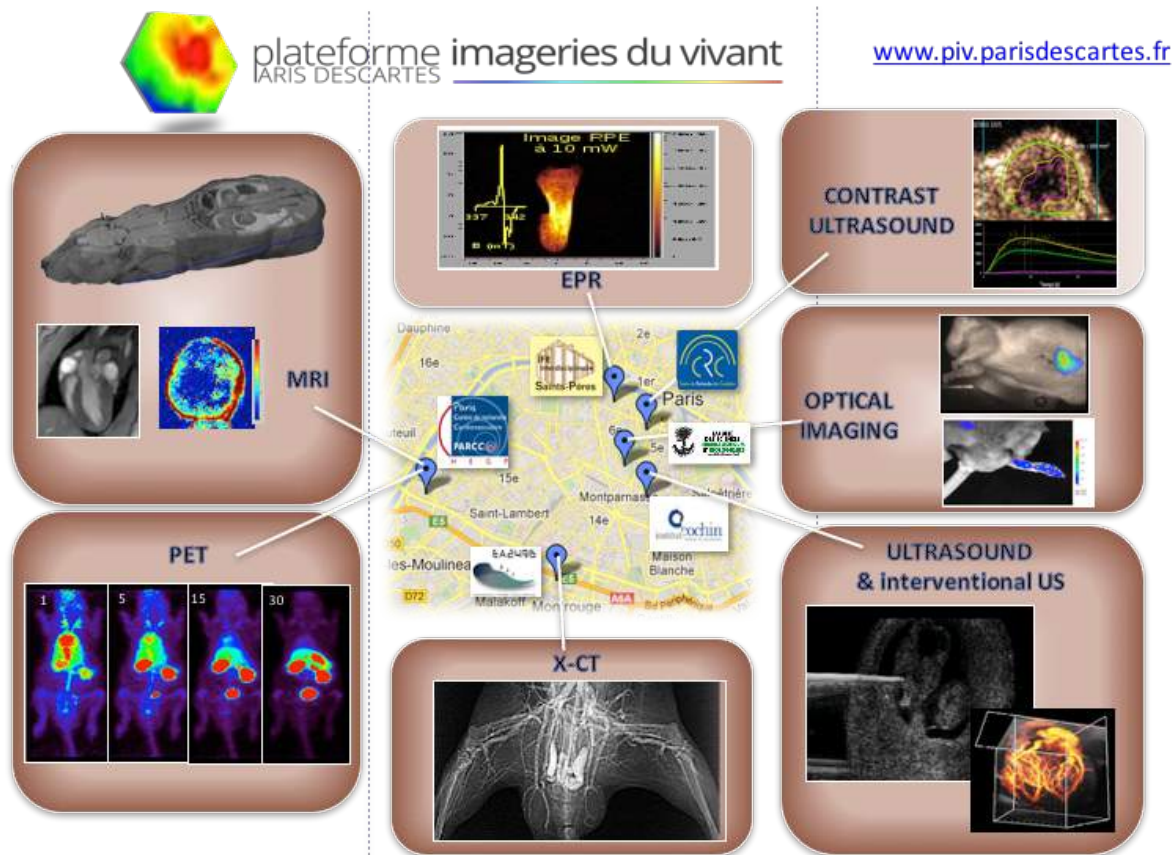
virtuelles, la mise en commun de licences logicielles, et l'élaboration d'un atlas annoté des données images disponibles sur le périmètre de SPC.

L'élaboration de cet atlas a été l'occasion de formaliser les éléments d'annotations requis pour une réutilisation des données, sous la forme d'une charte d'annotation évolutive et multidisciplinaire. Par ailleurs, Cloud IDV héberge notre solution de Crowd Sourcing, ainsi que des machines virtuelles qui sont des instances de serveurs de solutions commerciales de système de suivi de projets et de données.



La plateforme d'imagerie du vivant (PIV) de l'université Paris Descartes regroupe 9 laboratoires sur 7 sites du quart sud-ouest de Paris qui disposent d'équipements de pointe en : Echographie haute résolution et quantitative

- IRM préclinique et clinique
- Micro CT (scanner X)
- TEP-TDM
- Optique
- Résonance paramagnétique électronique (RPE)
- Traitement de données



Les équipes de la Plateforme d'Imageries du Vivant de Paris Descartes mettent leur expertise à la disposition de la communauté scientifique et technique pour étudier et réaliser tous projets de recherche et développement en imagerie.

CONTACT

- Site web : <http://piv.parisdescartes.fr>
- Email : imagerieanimale@medecine.univ-paris5.fr