



HEALTH
DATA HUB

Product Roadmap 2022

13 Avril 2022



Plateforme technologique du HDH

Offre actuelle

Les missions du HDH

- ❖ Mettre à disposition les **données** nécessaires aux projets visant l'intérêt public de manière **sécurisée** et **rapide**
- ❖ Optimiser leur traitement via un **panel d'outils à l'état de l'art** et, dans la mesure du possible, **open-source**

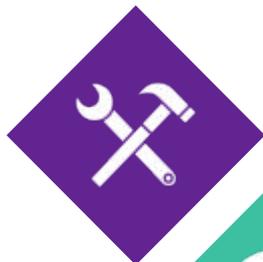
Offre technologique actuelle

	Langage de programmation et IDE	Infrastructure logicielle	Visualisation de données	Stockage de données	Imagerie	Collaboration
Outils disponibles						
Utilisation	<ul style="list-style-type: none">• Exploration des données• Machine learning• Traitement d'images	<ul style="list-style-type: none">• Requête et traitement de données en très grand volume	<ul style="list-style-type: none">• Visualisation des données	<ul style="list-style-type: none">• Organisation des données• Requête	<ul style="list-style-type: none">• Visualisation et annotation d'images	<ul style="list-style-type: none">• Suivi du développement• Versionning

Les thèmes de l'année 2022

Agrandir le panel d'outils pour traiter de nouveaux cas d'usages

- ❖ Calcul parallélisé (Cluster Spark)
- ❖ Viewer DICOM



Faciliter la mise à disposition des données

- ❖ Pré-traitement des données avant leur mise à disposition



Améliorer l'expérience utilisateur

- ❖ Gestionnaire de paquets (Python, R...)

Etre respectueux du RGDP

- ❖ Industrialisation de l'exercice des droits relatifs aux données personnelles

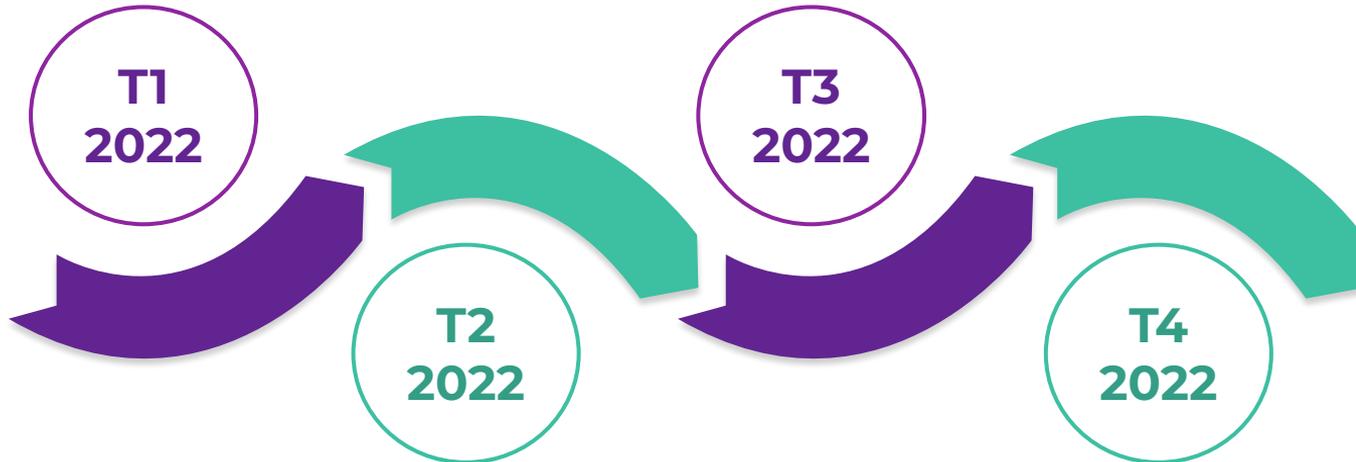
Simplifier l'opérabilité de la plateforme

- ❖ Industrialisation des déploiements des espaces utilisateurs

Roadmap Produit 2022

Gestion à l'ingestion des formats DICOM & JSON

Viewer DICOM
Pré-traitement des données avant mise à disposition



Gestionnaire de paquets (Python, R...)
Calcul parallélisé (Cluster Spark)
Industrialisation des déploiements des espaces utilisateurs

Portail d'authentification unique
R dans Jupyterlab
Industrialisation de l'exercice des droits

A blue-tinted photograph of several people's hands gathered around a table. One person is pointing at a tablet displaying a map. Other hands are holding pens and resting on notebooks. The scene suggests a collaborative meeting or workshop.

Principales features - T1 2022

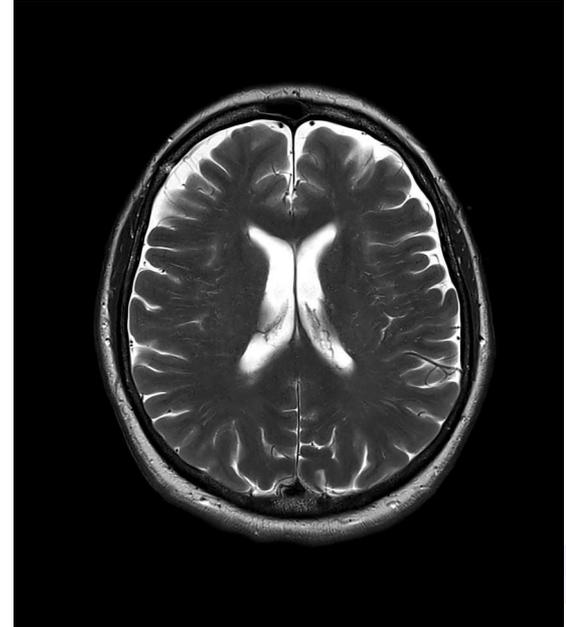
Gestion à l'ingestion des formats DICOM & JSON

Contexte

- ❖ Le **format d'imagerie DICOM** et le **format JSON** sont régulièrement utilisés en recherche médicale
- ❖ Certains projets utiliseront ces formats afin de réaliser leur étude sur la plateforme du HDH

Nouveauté

- ❖ Les données au format DICOM et JSON pourront être **ingérées** dans la plateforme technologique du HDH





Principales features - T2 2022

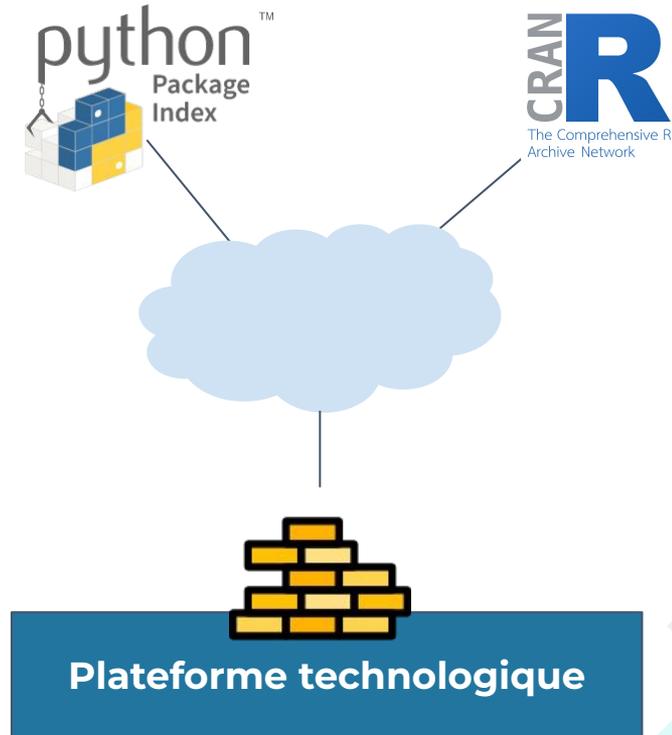
Gestionnaire de paquets (Python, R...)

Contexte

- ❖ Les projets demandent l'installation de paquets (outils de traitement de données) répondant à leurs besoins. Aujourd'hui **plus de 700 paquets sont installés**
- ❖ La gestion des demandes, validation et installation de ces paquets requiert beaucoup d'actions manuelles

Nouveauté

- ❖ Un **outil dédié, automatisant ce processus** et offrant plus de **flexibilité** aux projets dans le choix et l'installation de leurs paquets sera mis à disposition sur la plateforme



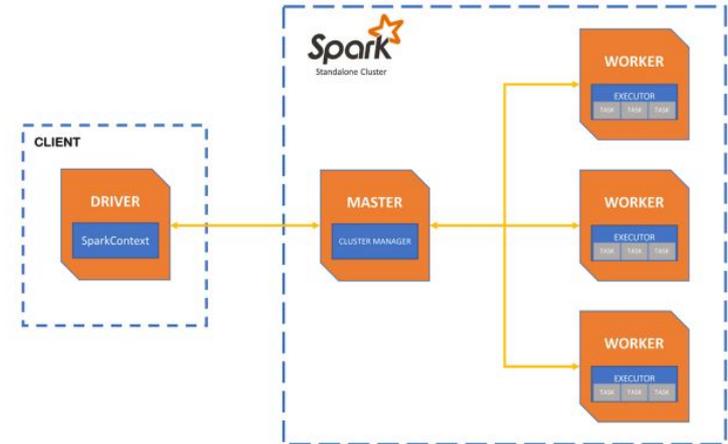
Calcul parallélisé (Cluster Spark)

Contexte

- ❖ La **volumétrie** des données traitée par certains projets peut atteindre **plusieurs milliards de lignes**
- ❖ Le traitement ou le requêtage de ces données avec des solutions de traitement séquentiel peut **s'étaler sur des heures/jours**

Nouveauté

- ❖ **Spark**, en tant qu'outil de référence pour le **traitement parallélisé des données**, sera mis à disposition des utilisateurs



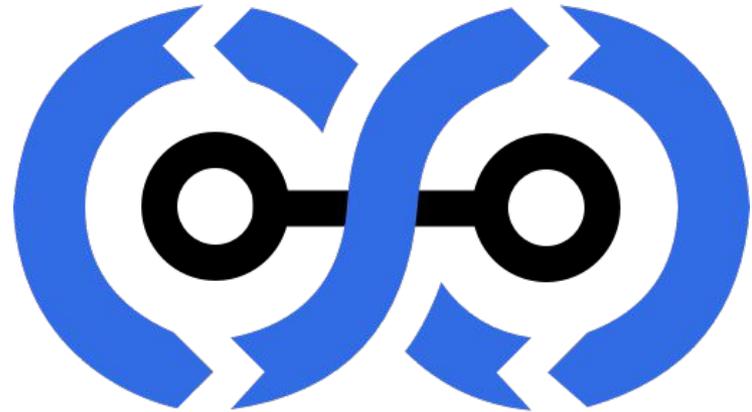
Industrialisation des déploiements d'espaces

Contexte

- ❖ **Chaque projet** est réalisé dans un **espace qui lui est dédié** et créé spécifiquement pour le projet
- ❖ Chaque déploiement ou mise à jour d'espace demande actuellement une **part d'actions manuelles** par un opérateur du HDH

Nouveauté

- ❖ Des **solutions d'automatisation et d'industrialisation des déploiements** et des mises à jour des espaces seront proposées



A blue-tinted photograph showing several people's hands gathered around a table. One person is pointing at a tablet displaying a bar chart. There are notebooks and pens on the table. A large white triangle is overlaid on the left side of the image.

Principales features - T3 2022

Viewer DICOM

Contexte

- ❖ Le format **DICOM** est le format standard en imagerie médicale. Le traitement des fichiers DICOM peut être fait sur la plateforme en utilisant Python ou R.
- ❖ Les profils médecins ont besoin de pouvoir :
 - effectuer des segmentations
 - visualiser des contrôles en sortie des algorithmes de deep learning

Nouveauté

- ❖ Un **viewer DICOM** sera mis à disposition des utilisateurs



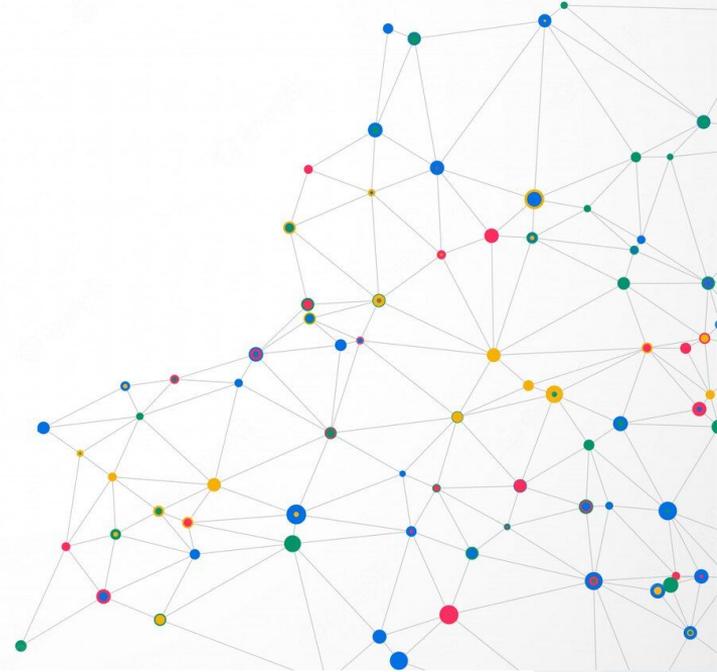
Pré-traitement des données avant mise à disposition

Contexte

- ❖ Aujourd'hui les référents données du HDH ne peuvent préparer les données que de manière dédiée à un projet
- ❖ Il leur sera utile de **pré-traiter les données**, et créer des sous-produits de données **indépendamment d'un projet, afin de pouvoir les réutiliser** pour tous

Nouveauté

- ❖ Un **nouvel espace de préparation non spécifique à un projet** sera mis à disposition des référents données du HDH





Principales features - T4 2022

Portail d'authentification unique

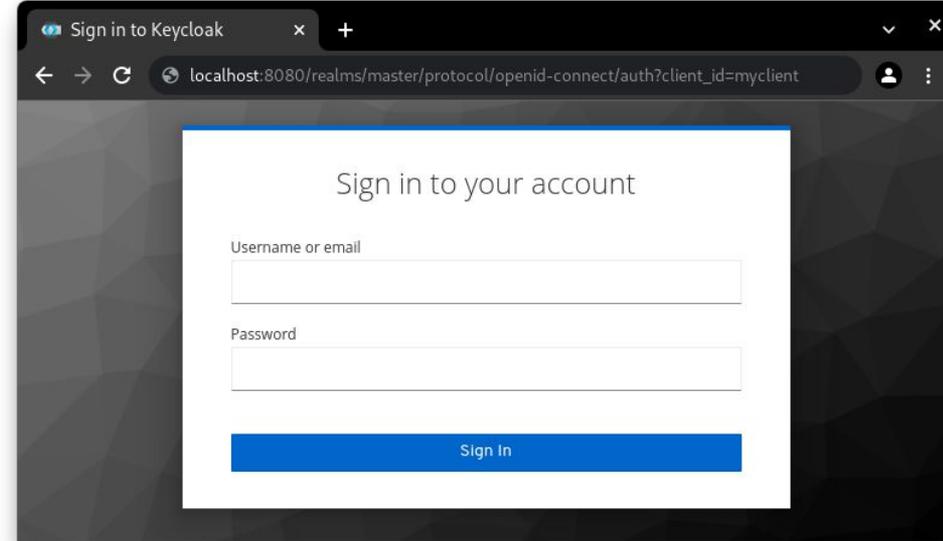
Contexte

Actuellement, les utilisateurs se **connectent aux applications de leur espace** via :

- ❖ l'interface d'authentification de GitLab pour JupyterLab et GitLab. Les utilisateurs s'authentifient une fois via cette interface et peuvent ensuite accéder aux deux outils
- ❖ l'interface de connexion de chaque outil pour les autres outils

Nouveauté

- ❖ Un **portail d'authentification unique** sera mis en place afin de permettre aux utilisateurs de se connecter une fois et d'accéder ensuite à **toutes les applications**



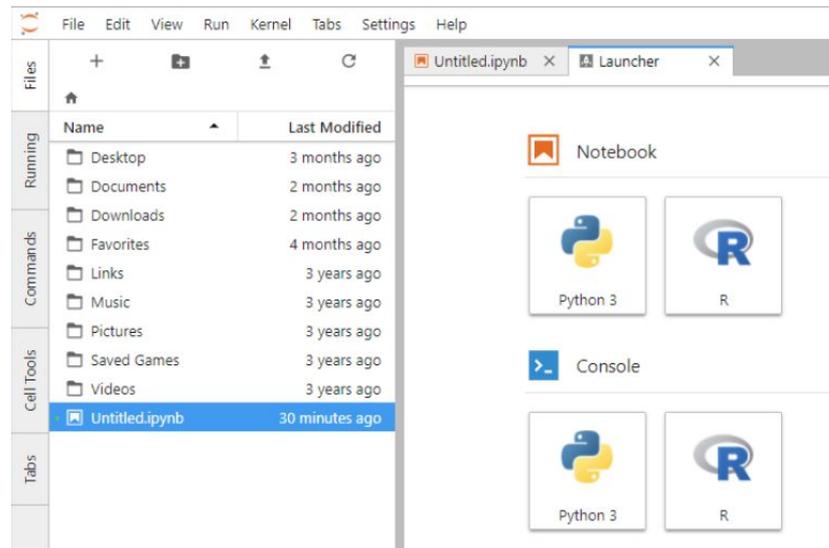
R dans JupyterLab

Contexte

- ❖ **R** est actuellement l'un des langages les plus utilisés dans le traitement des données, particulièrement dans la recherche biomédicale
- ❖ La plateforme technologique propose actuellement l'outil **RStudio** pour permettre aux projets d'utiliser R

Nouveauté

- ❖ Il sera désormais possible d'**utiliser R dans JupyterLab**, permettant aux utilisateurs travaillant sur Python et R de rester sur un même IDE



Exercice des droits relatifs aux données personnelles

Contexte

- ❖ Conformément au **RGPD** et au code de la santé publique, tout individu doit pouvoir **exercer ses droits relatifs aux données personnelles** (ex : droit d'opposition)

Nouveauté

- ❖ Un **système de recueil et de traitement de demandes** sera mis en place, à travers un **formulaire web** accessible aux individus qui souhaitent exercer leurs droits relatifs aux données personnelles





Suivez-nous sur les réseaux sociaux !

