



Gestion et exploitation de données COVID

Tiphaine Diot, Olivier Bernard, Pierre Croisille, Frédéric Cervenansky

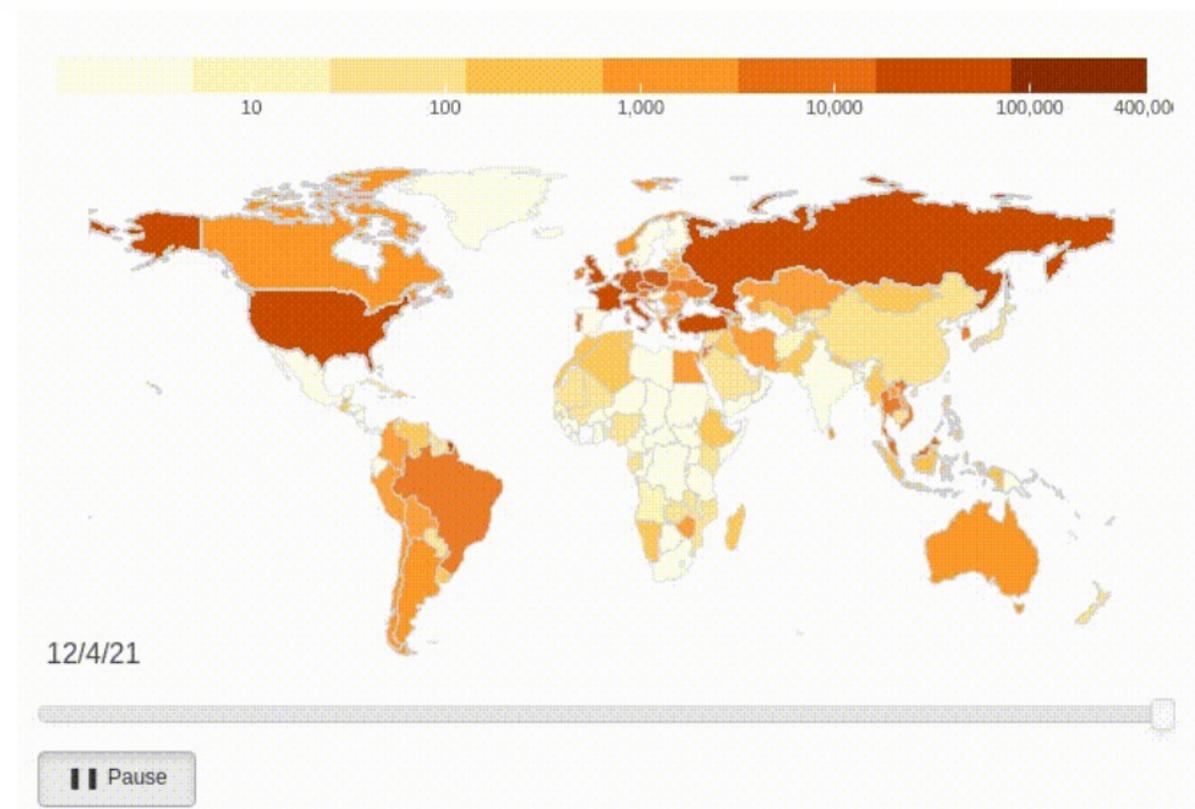
Journées Calcul et Données 2021

Contexte: crise mondiale

Début projet Feb. 2020 : CHU St-Etienne décide de collecter des données d'imagerie et des données cliniques en suivant un protocole clairement identifié

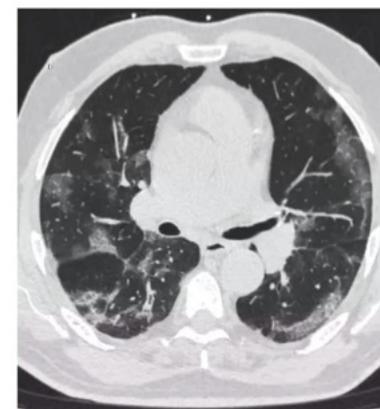
Daily Confirmed New Cases

Source: John Hopkins University, coronavirus Resource center

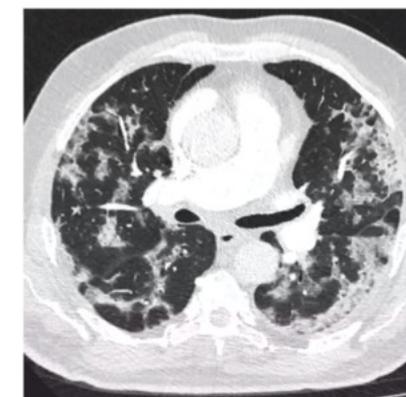


Contexte : Imagerie CT

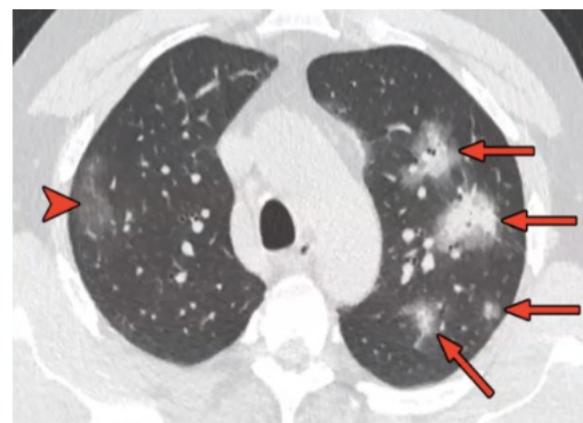
- Imagerie CT apporte des **informations pertinentes & complémentaires** aux données PCR
- **Signatures caractéristiques du COVID-19**
 - ↳ Verre Dépoli (Ground Glass Opacity - GGO)
 - ↳ Pavage fou (Crazy Paving)
 - ↳ Signes d'halo (Halo Sign)
 - ↳ Dilatation vasculaire
 - ↳ Evolution temporelle
 - ↳ ...



Premier scan CT



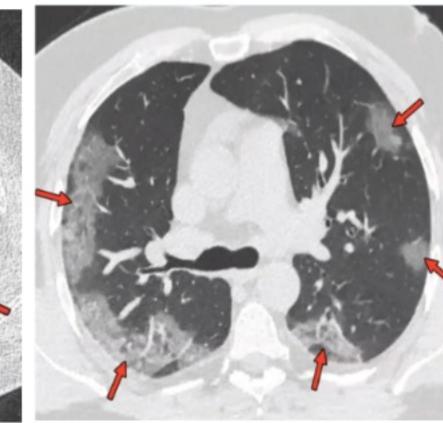
Après 10/15 jours



Signe d'halo



Pavage fou



Verre depoli

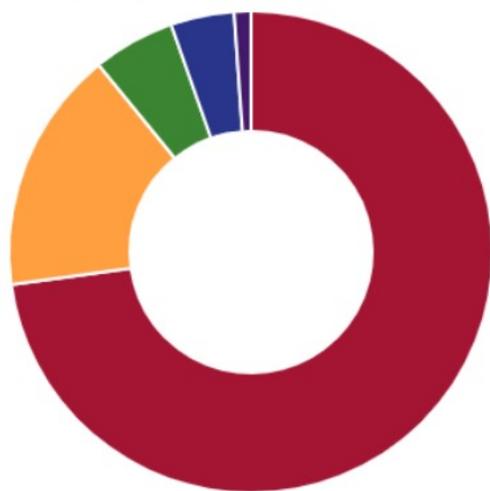
COVID-19 lung CT lesion segmentation challenge: covid-segmentation.grand-challenge.org

Contexte : base de données images

Images Total: 1157 séquences d'imagerie CT pour 800 patients (> 2 To)
Plusieurs protocoles d'acquisition (injecté ou non, ventilation). Conversion automatique d'images (DICOM, Nifti & mhd).

Patient Distribution number of exams

1 Session 2 Sessions 3 Sessions
4 Sessions 5 Sessions



Patient distribution of protocol variants

S SI SIV



Proportion of Patient with COVID

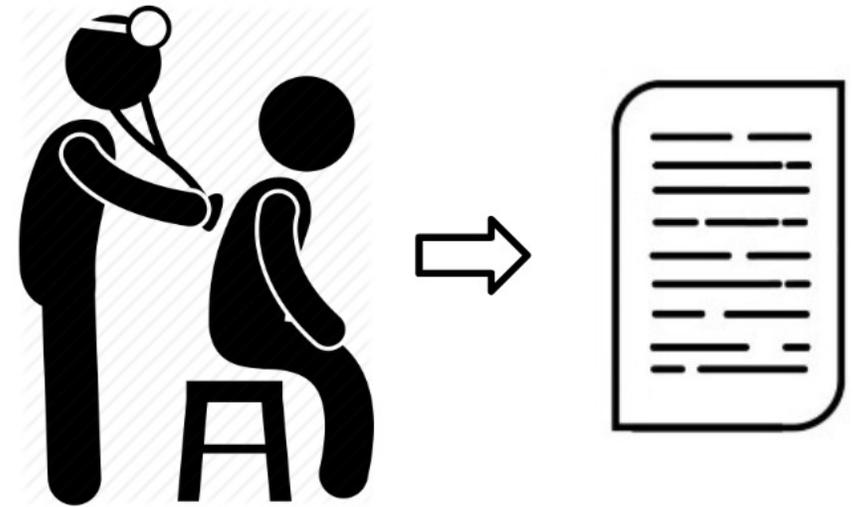
Confirmed COVID-19
Not Confirmed COVID-19



Contexte : données cliniques associées

Electronic Clinical Reports / Métadonnées pour chaque patient: très sensibles

- Suivi sur 1 mois
 - ↳ Etude longitudinale possible sur une partie de la base
- Jusqu'à 400 champs par patient
 - ↳ Historique médical
 - ↳ Signes vitaux & biologie à l'admission
 - ↳ Résultats XRay/CT/US des poumons
 - ↳ Issue à J30
- Données fortement hétérogènes



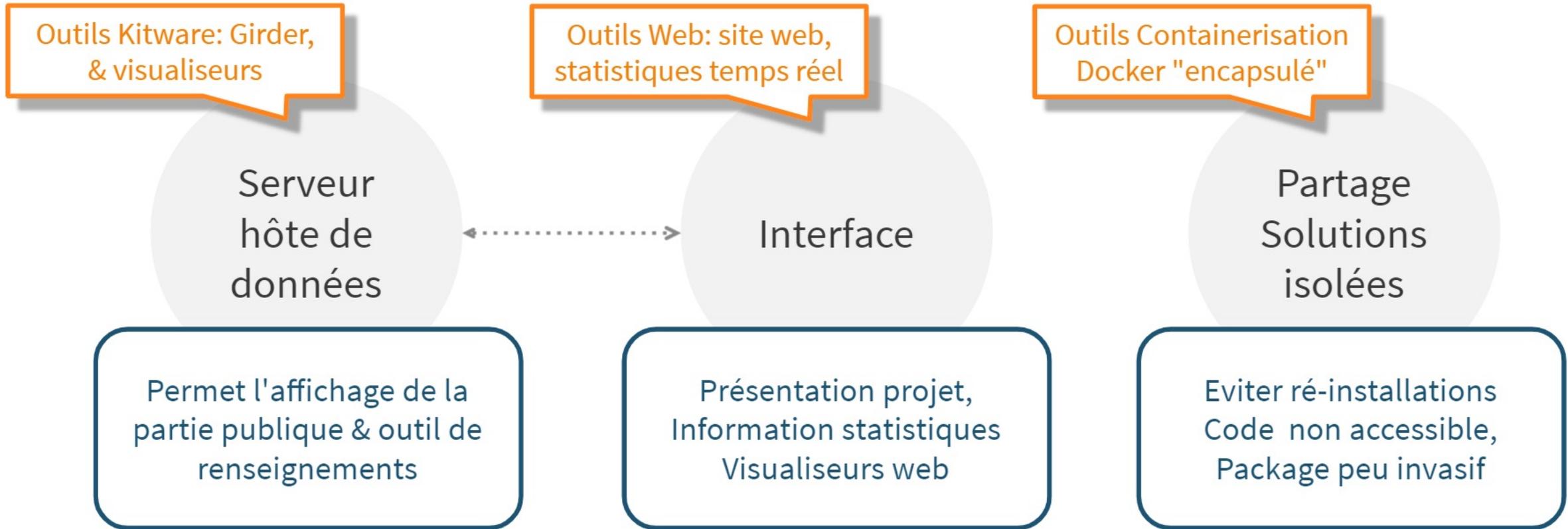
ClinicalTrials.gov Identifier: NCT04377685

Contexte : problématique

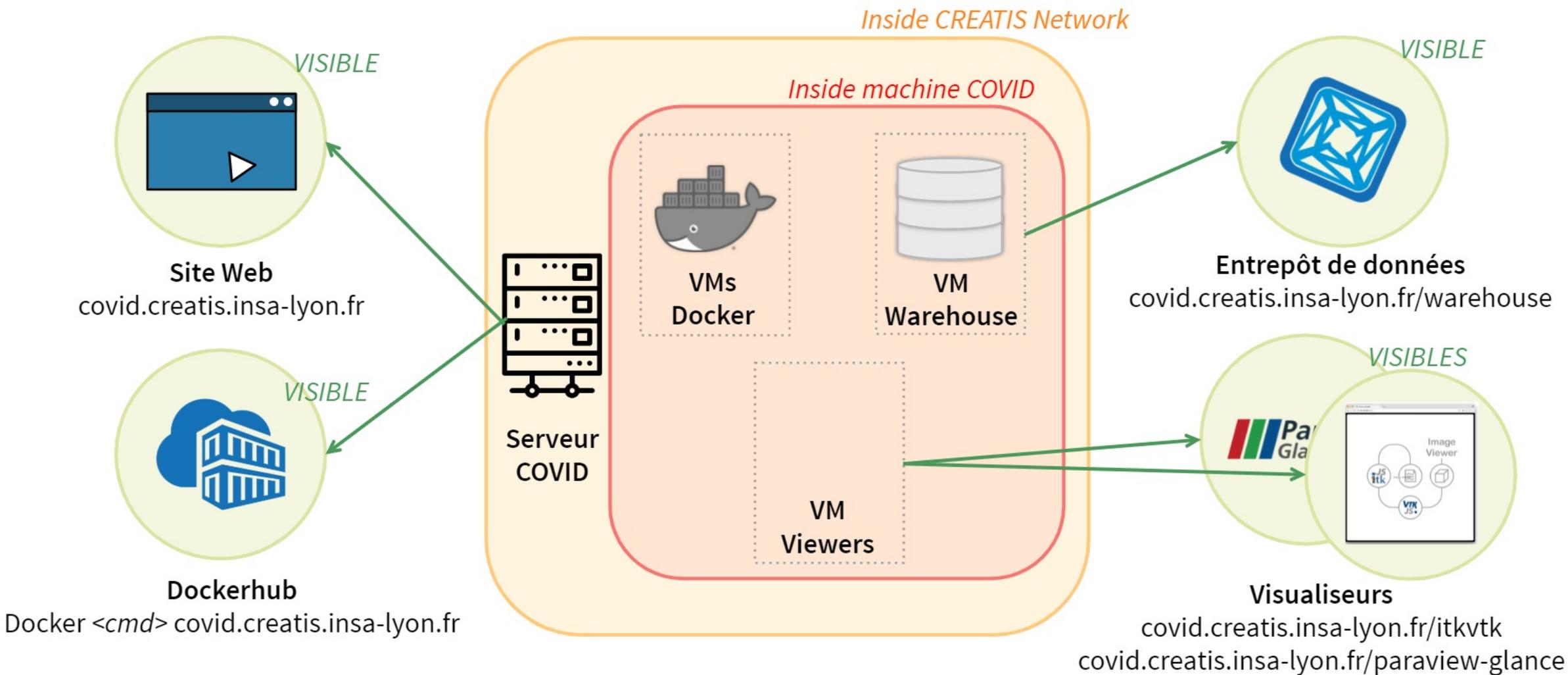
Fournir à la communauté scientifique française publique
des services dédiés à l'analyse de données COVID



Infrastructure: 3 services

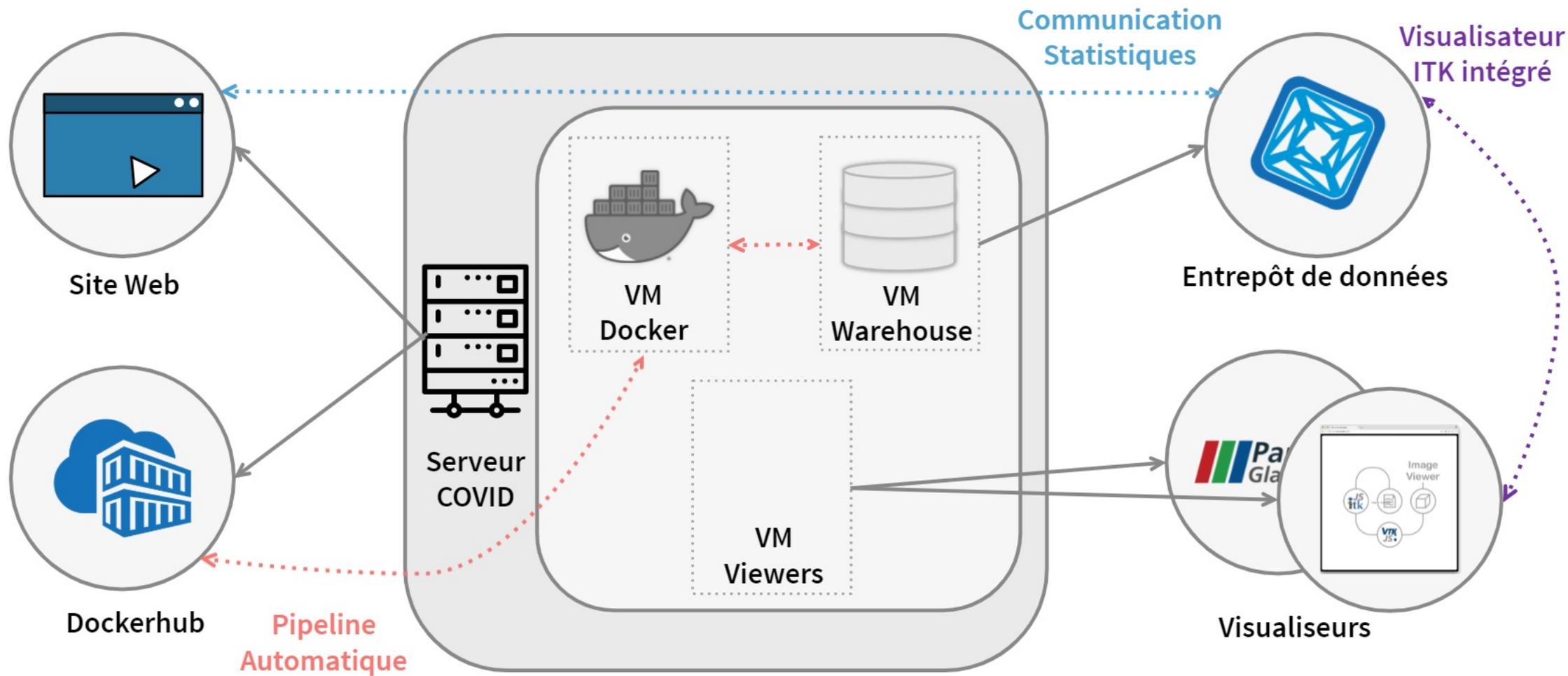


Infrastructure: général



- I.
- II.
- III.
- IV.

Infrastructure: général



I.
II.
III.
IV.

Infrastructure: entrepôt de données

Outil Girder: "plateforme personnalisable de gestion de données qui fournit un stockage de données en nuage avec une interface web pour vos données et vos fichiers."

- Enregistrement obligatoire, Compte validé par admins
- + date expiration de compte (1 an)

Accès

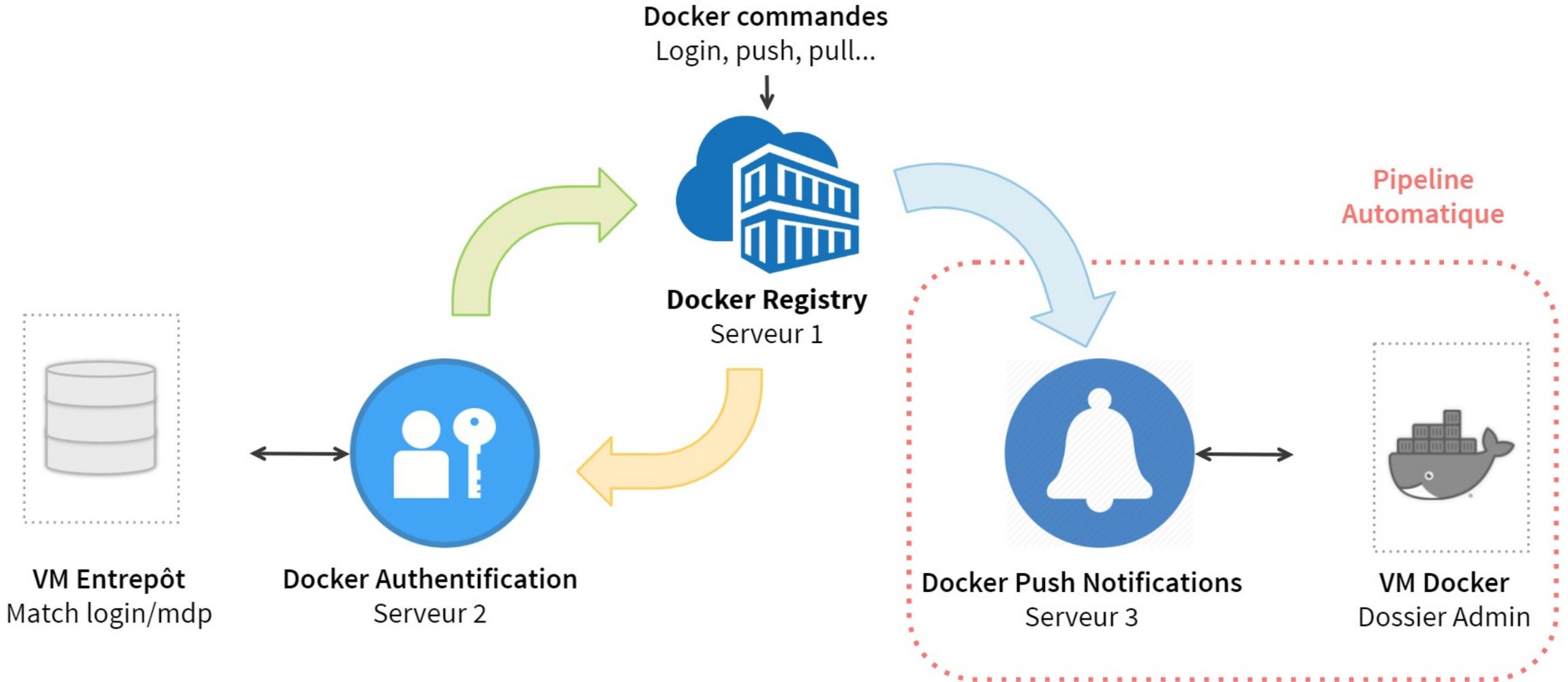
- Outils pipeline utilisateur (windows & linux)
- 2 Collections
 - Privée : base complète
 - Publique : Images & metadonnées décorrelées

Architecture

- Outil stats métadonnées
- Visualiseur intégré
- Journal des soumissions
- Système mails (admins)
- Outil conversion DICOM
- ...

Ajout de fonctionnalités

Infrastructure: Dockerhub

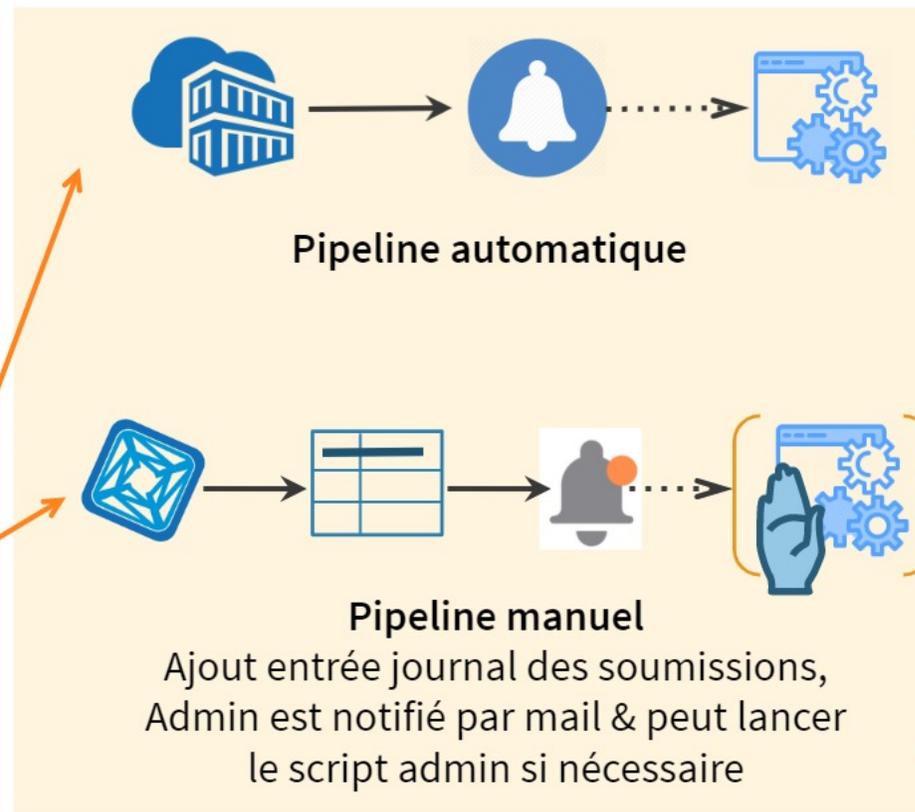


- I.
- II.
- III.
- IV.

Infrastructure: pipeline

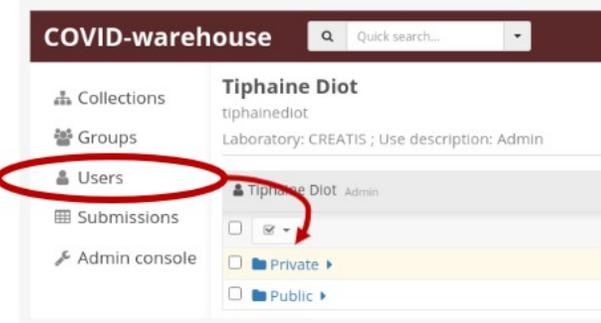
1. Téléchargement dossier utilisateur
2. Création de l'image docker
3. Partage de l'image solution

Côté Utilisateur



Côté Administrateur : 2 modes

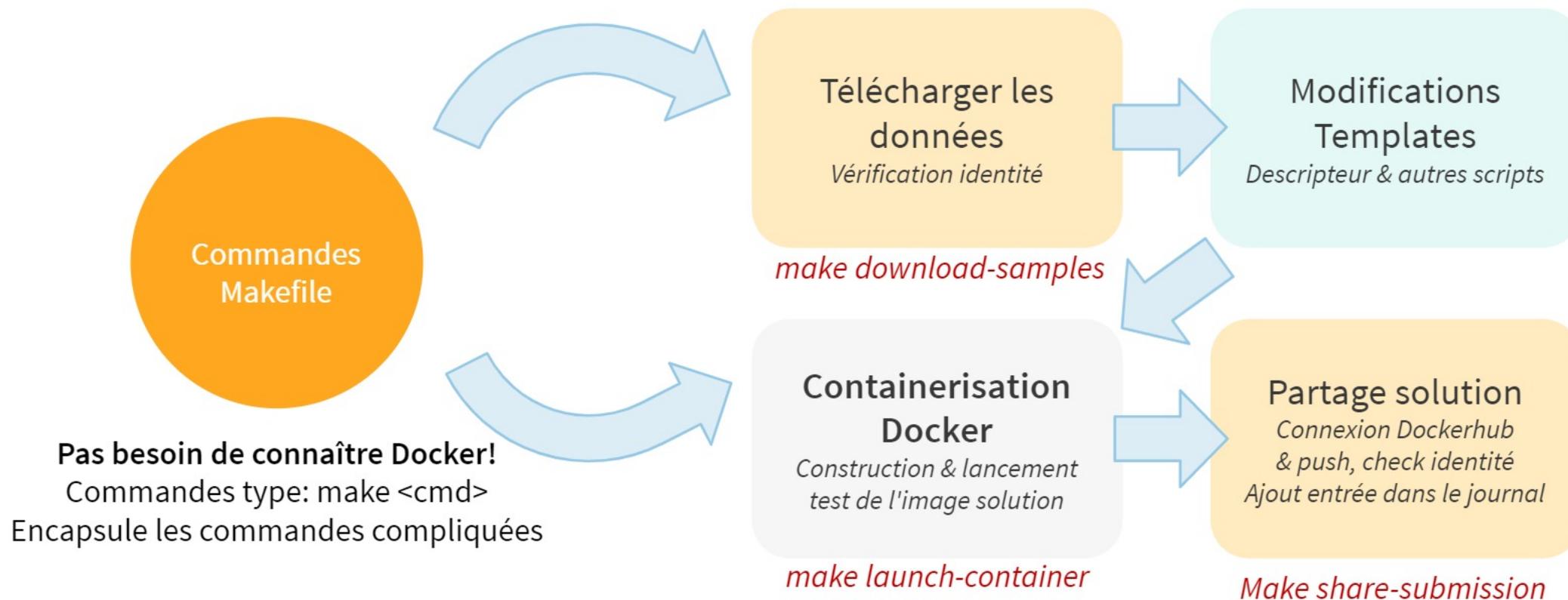
4. Notifier quand le statut de l'image est mis à jour
5. Résultats en ligne sur l'entrepôt



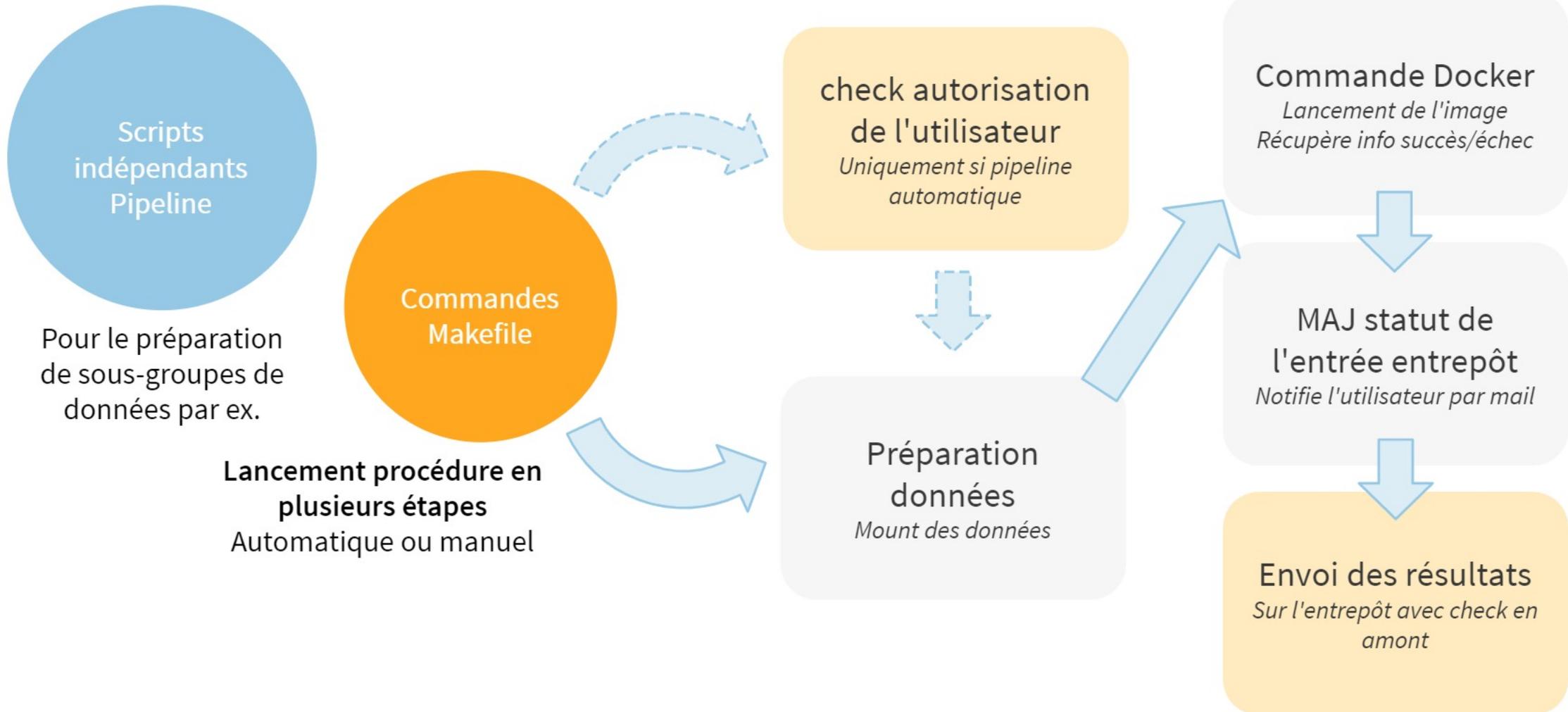
Côté Utilisateur

I.
II.
III.
IV.

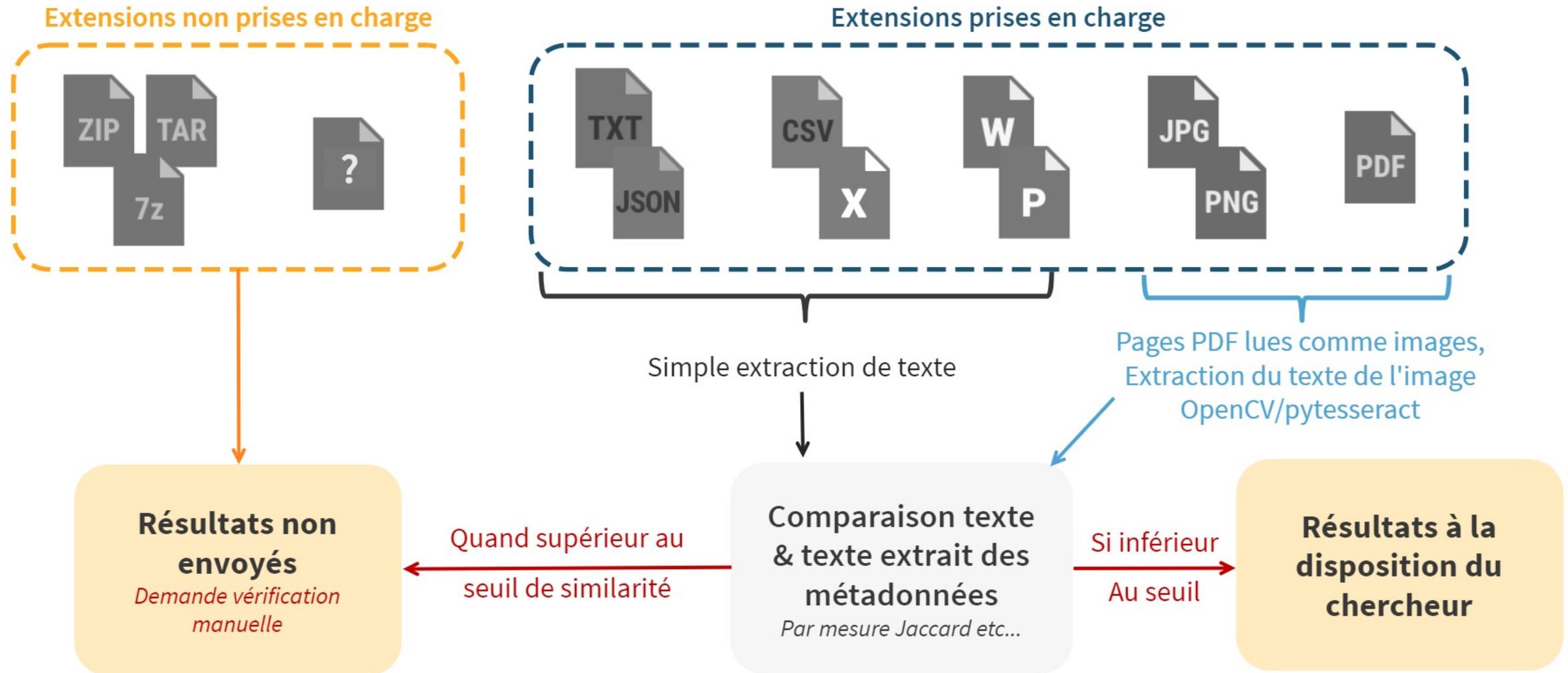
Infrastructure: dossier utilisateur



Infrastructure: dossier administrateur

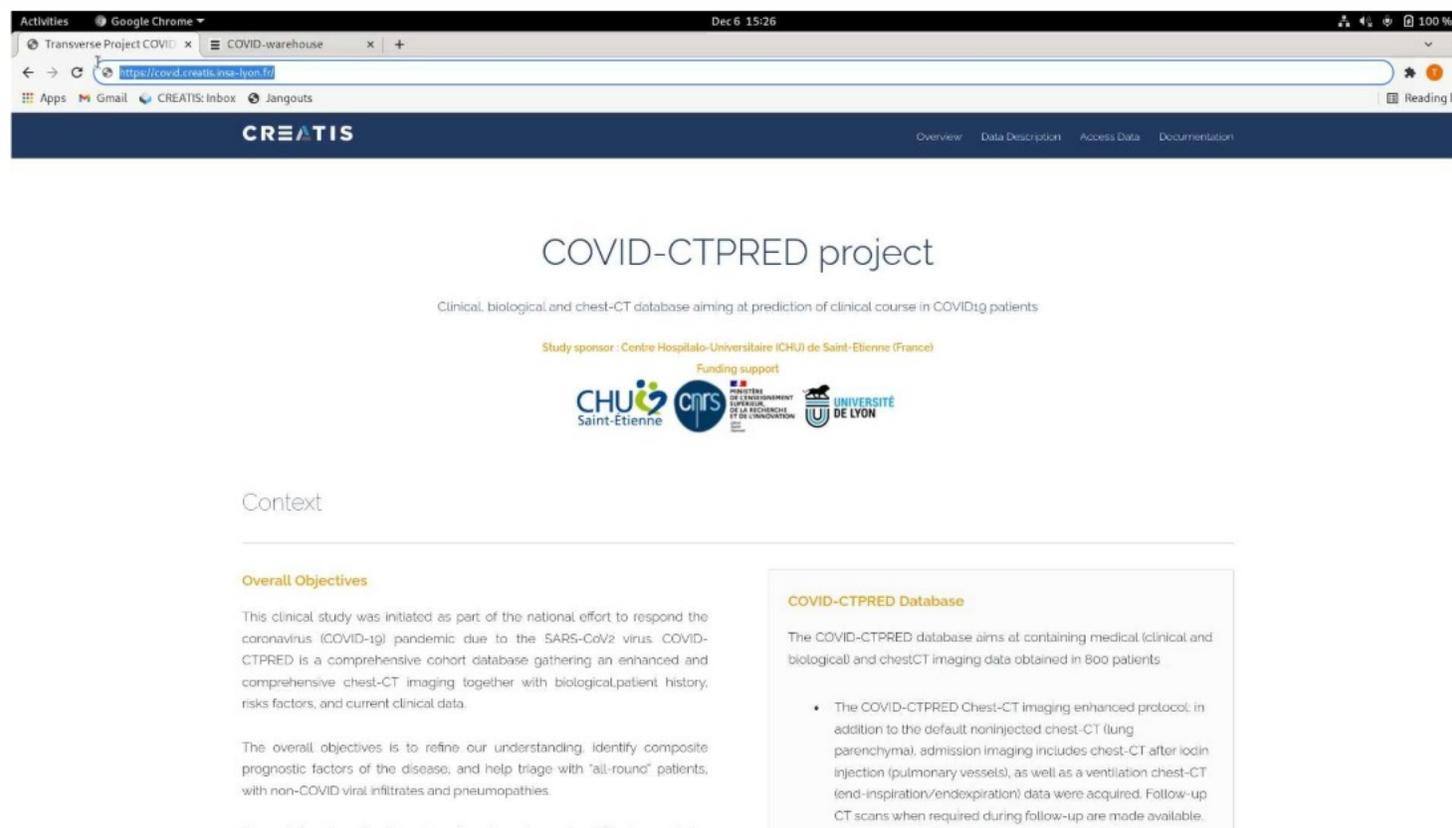


Infrastructure: vérification résultats



Démonstration: Vidéo

Rappel URL: <https://covid.creatis.insa-lyon.fr/>



The screenshot shows a web browser displaying the COVID-CTPRED project page. The browser's address bar shows the URL <https://covid.creatis.insa-lyon.fr/>. The website header features the CREATIS logo and navigation links for Overview, Data Description, Access Data, and Documentation. The main content area is titled "COVID-CTPRED project" and describes it as a "Clinical, biological and chest-CT database aiming at prediction of clinical course in COVID-19 patients". It lists the study sponsor as "Centre Hospitalo-Universitaire (CHU) de Saint-Etienne (France)" and funding support from "CHU Saint-Etienne", "CNRS", "MINISTÈRE DE L'ÉTABLISSEMENT SUPÉRIEUR DE LA RECHERCHE ET DE L'INNOVATION", and "UNIVERSITÉ DE LYON". Below this, a "Context" section is visible, with a sub-section for "Overall Objectives" and a "COVID-CTPRED Database" section.

Activities Google Chrome Dec 6 15:26

Transverse Project COVID-19 COVID-warehouse

<https://covid.creatis.insa-lyon.fr/>

Apps Gmail CREATIS: Inbox Jangouts Reading list

CREATIS Overview Data Description Access Data Documentation

COVID-CTPRED project

Clinical, biological and chest-CT database aiming at prediction of clinical course in COVID-19 patients

Study sponsor : Centre Hospitalo-Universitaire (CHU) de Saint-Etienne (France)

Funding support

CHU Saint-Etienne CNRS MINISTÈRE DE L'ÉTABLISSEMENT SUPÉRIEUR DE LA RECHERCHE ET DE L'INNOVATION UNIVERSITÉ DE LYON

Context

Overall Objectives

This clinical study was initiated as part of the national effort to respond to the coronavirus (COVID-19) pandemic due to the SARS-CoV-2 virus. COVID-CTPRED is a comprehensive cohort database gathering an enhanced and comprehensive chest-CT imaging together with biological, patient history, risks factors, and current clinical data.

The overall objectives is to refine our understanding, identify composite prognostic factors of the disease, and help triage with 'all-round' patients, with non-COVID viral infiltrates and pneumopathies.

COVID-CTPRED Database

The COVID-CTPRED database aims at containing medical (clinical and biological) and chest-CT imaging data obtained in 800 patients

- The COVID-CTPRED Chest-CT imaging enhanced protocol: in addition to the default noninjected chest-CT (lung parenchyma), admission imaging includes chest-CT after iodine injection (pulmonary vessels), as well as a ventilation chest-CT (end-inspiration/end-expiration) data were acquired. Follow-up CT scans when required during follow-up are made available.

Discussion & conclusions

Réponses aux problématiques

- **Contrôle de l'accès de l'entrepôt** : compte nécessaire
- Demande **identifiants pour envoyer une solution** sur le dockerhub

Accès & Données

- Les infrastructures accessibles uniquement via clé ssh admins
- Pipeline automatique protégé par un mot de passe

Infrastructure protégée

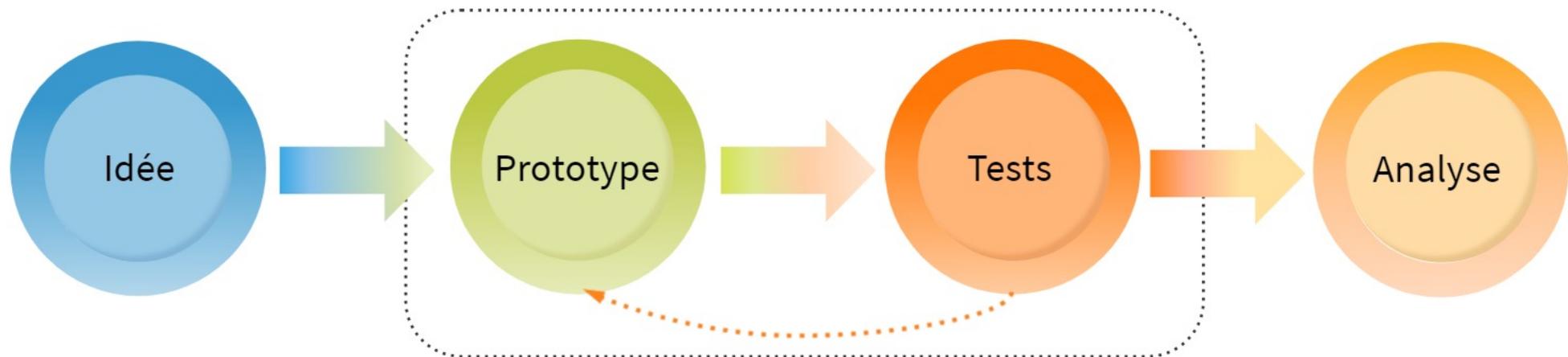
- Pipeline automatique ouvert aux **utilisateurs de confiance**
- Avant envoi, **vérification si résultats similaires aux métadonnées pour éviter fuite**

Pipeline & résultats

Discussion & conclusions

Suite du Projet

- Ajout d'une page sur l'entrepôt de données permettant aux utilisateurs de choisir des "options"
- Amélioration itérative
 - ↳ Intégration continue des futures propositions & avis utilisateurs





Merci pour votre attention !

Des questions ?

