

I-SITE CAP 20-25

Programme DATA - Projet 2022-2027

Tout en s'appuyant sur les multiples acquis de la période 2017-2020, la deuxième phase du programme « INSTRUMENTS » s'inscrit dans un double objectif :

- Poursuivre la vocation initiale de ce programme de jouer un rôle de support unissant les compétences du site Clermontois pour lever les verrous scientifiques et technologiques rencontrés dans la réalisation des enjeux liés aux futurs CIRs,
- Pérenniser le programme « INSTRUMENTS » via son intégration dans la stratégie scientifique du site Clermontois. L'objectif est d'opérer une ouverture thématique du programme « INSTRUMENTS » en lui donnant la possibilité d'adresser des questions fondamentales ou applicatives prioritaires par l'établissement et qui se situent dans le périmètre thématique du programme.

Pour atteindre ces objectifs, deux évolutions importantes du programme sont proposées : d'une part, un élargissement thématique et applicatif et d'autre part une évolution de sa structuration et de son pilotage.

La thématique générale du futur programme, dont l'intitulé proposé est DATA (Données : Acquisition, Transfert, Apprentissage), s'articule autour de la « donnée », cette dernière étant vue comme l'élément fédérateur de l'ensemble des activités de ce programme. Il s'agit de fédérer les compétences du site Clermontois permettant d'adresser le **cycle de vie complet de la « donnée »** :

- De la mesure et de l'acquisition des données via des dispositifs variés,
- Au transfert des données via des réseaux de capteurs,
- Au stockage, l'intégration et la gestion des données à l'aide de techniques modernes du bigdata,
- Et enfin à la valorisation des données par des techniques de fouille de données et d'intelligence artificielle.

Le programme DATA sera structuré autour de trois pôles :

- IoT : mesures, acquisition et transfert des données via les réseaux de capteurs.
- BigData : stockage, intégration et gestion de grandes masses de données.
- IA : fouille de données et Intelligence Artificielle.

La deuxième évolution du programme « INSTRUMENTS » est de nature structurelle. Compte tenu des forces en présence sur le site Clermontois, de la forte intersection entre les périmètres scientifiques et les communautés cibles du programme « INSTRUMENTS » d'une part, et de la thématique émergente IA d'autre part, il est proposé l'intégration de la thématique émergente IA de la politique de site au sein du futur programme DATA. Cela permettra à la fois de développer les aspects théoriques de l'IA et d'irriguer les domaines applicatifs des futurs CIRs de l'ISITE et d'autres domaines d'application, en synergie avec la dynamique impulsée par la Région AURA sur le domaine de l'IA.

Le programme adressera donc, comme précédemment, les verrous technologiques rencontrés dans la réalisation des enjeux liés aux CIRs, consolidant et poursuivant les actions déjà réalisées et en cours sur les challenges. Ceci concerne le passage à l'échelle du CLOUD CEBA du Challenge 1, la valorisation des briques technologiques validées dans ce cadre (en particulier au sein de l'EQUIPEX TERRA-FORMA). Le projet EMOB n'en est qu'à sa phase 1 qui doit se terminer en début 2022. La phase 2 impliquera le programme DATA sur le prototype d'équipement pour mesurer la mobilité et des aspects gestion/entrepôts de données. En parallèle, il sera nécessaire de concevoir avec les acteurs de la santé les algorithmes d'exploitation scientifique devant en résulter. Les besoins dans le domaine des risques volcaniques (CH4) sont immenses. Ils vont du besoin de définir de nouvelles approches pour collecter les informations sur des sites difficiles d'accès, la nécessité de nouveaux types de mesures (le radon dans les panaches volcaniques par exemple). Par ailleurs, le suivi temporel fin (inexistant actuellement), et la capacité d'analyse 'online' pour orienter de manière dynamique les mesures sont des enjeux d'importance pour ce domaine.

Tous ces champs sont déjà en réflexion dans le programme DATA, en synergie avec le nouveau programme IDEAL (de l'Intelligence Distribuée A L'Apprentissage automatique : approches innovantes pour la gestion des données massives de la recherche) du CPER2021-25, co-construit avec le programme « INSTRUMENTS ».

En plus de la consolidation des interactions avec les CIRs, les domaines de la santé (hors mobilité) et des SHS ont été identifiés comme des cibles potentielles pour les actions du programme DATA.

La structuration du programme DATA est présentée dans la figure 1. Le pilotage sera assuré par un COPIE représentatif des forces en présence, des expertises existant dans chacun des trois thèmes de l'IOT, du bigdata et de l'IA (3 représentants par thème), tout en assurant une continuité dans les actions déjà en cours.

Composition du COPIE :

- | | | | |
|----------------------|-------------|--------------------|-----------------|
| - Vincent Barra | ISIMA-LIMOS | - Antoine Mahul | DOSI-MESOCENTRE |
| - Emmanuel Bergeret | IUT-LPC | - Dominique Pallin | LPC |
| - Jean-Pierre Chanet | INRAe | - Céline Teulière | IP |
| - Arnaud Guillin | LMBP | - Farouk Toumani | ISIMA-LIMOS |
| - Oussama Habachi | IUT-LIMOS | | |

Ce COPIE pourra se réunir en formation élargie, en incluant des membres du collège des usagers (issus des futurs CIRs, ou de domaines applicatifs que le programme a vocation d'adresser, comme le domaine de la santé hors CIR, ou SHS)

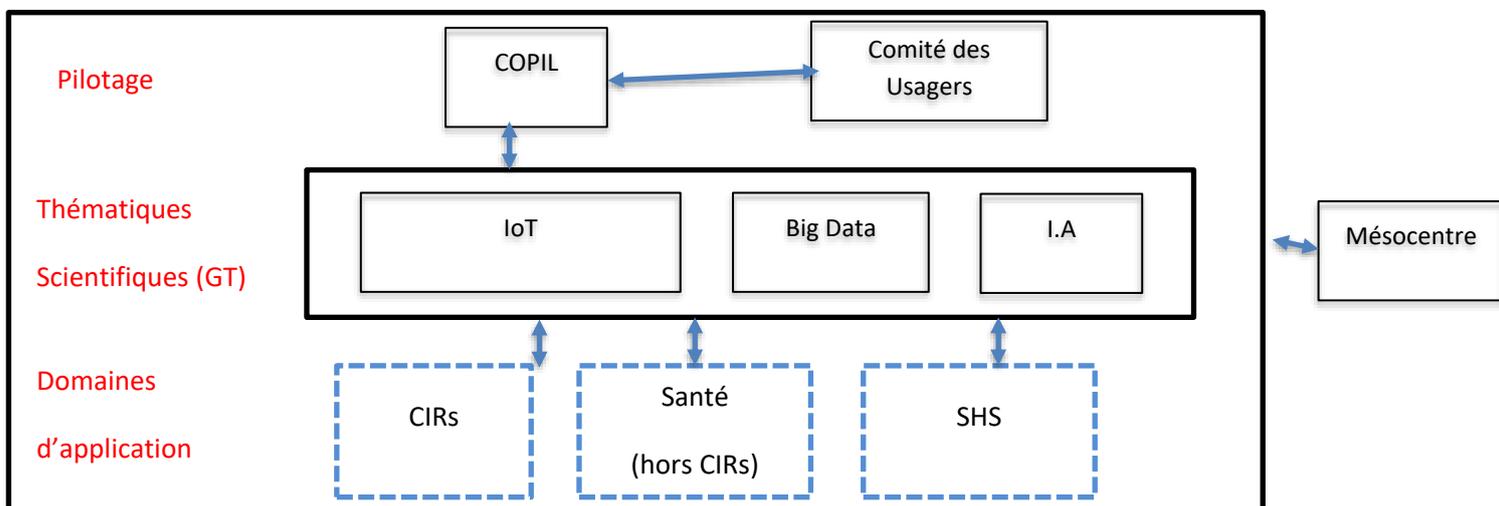


Figure 1 – Organisation du programme instruments