



Activités IA ODATIS / Data Terra

Dimitry Khvorostyanov – Resp. Innovation IA ODATIS

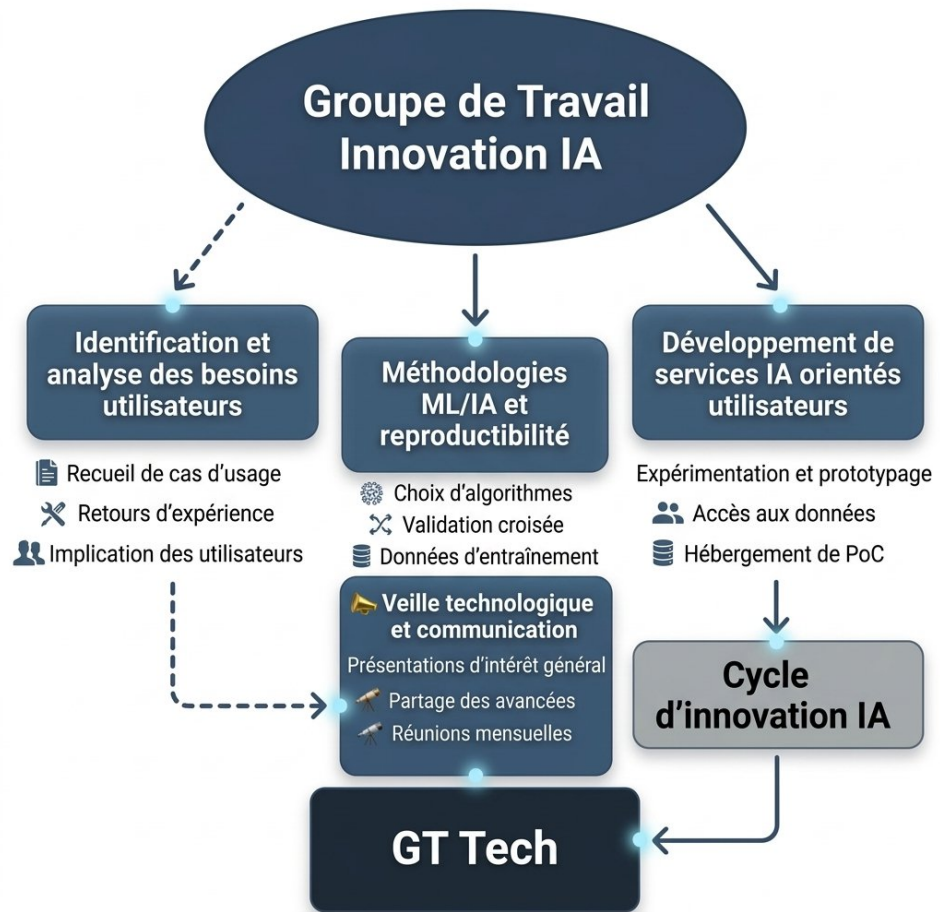
Journée ODATIS Océan Indien – 18/09/2026



Plan de présentation

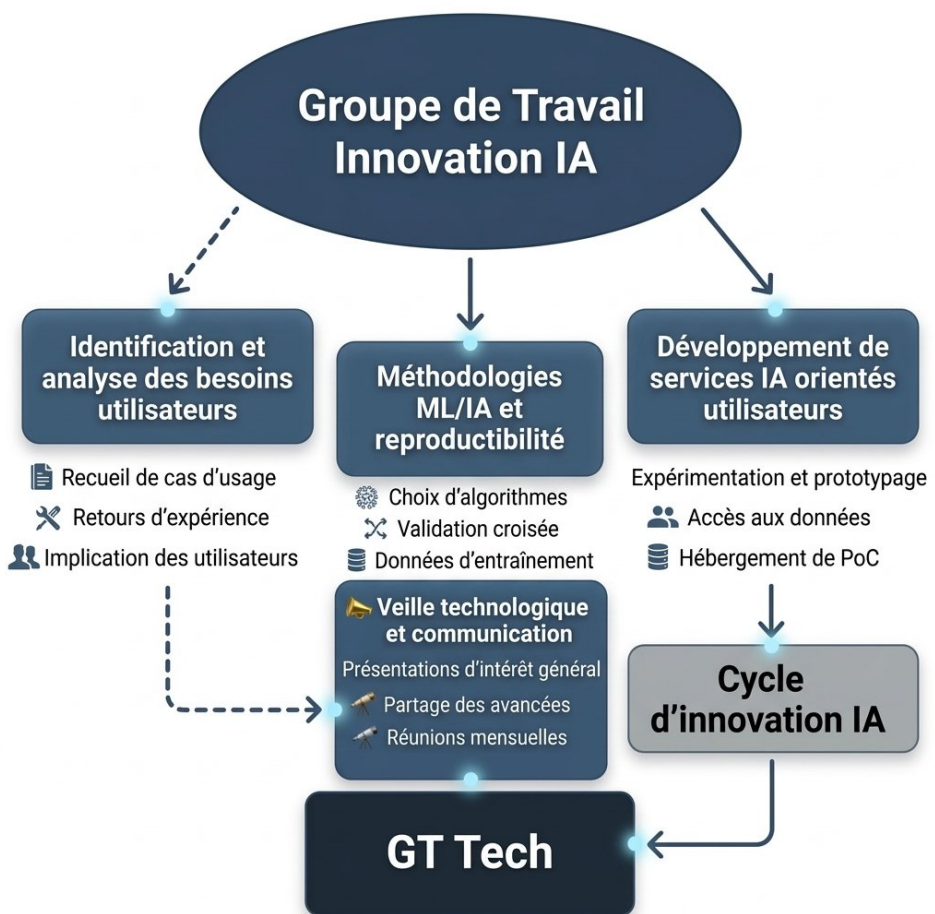
1. Fonctionnement du GT Innovation IA ODATIS
2. Des activités en cours
3. Projets récents Data Terra et Infrastructures

GT Innovation IA

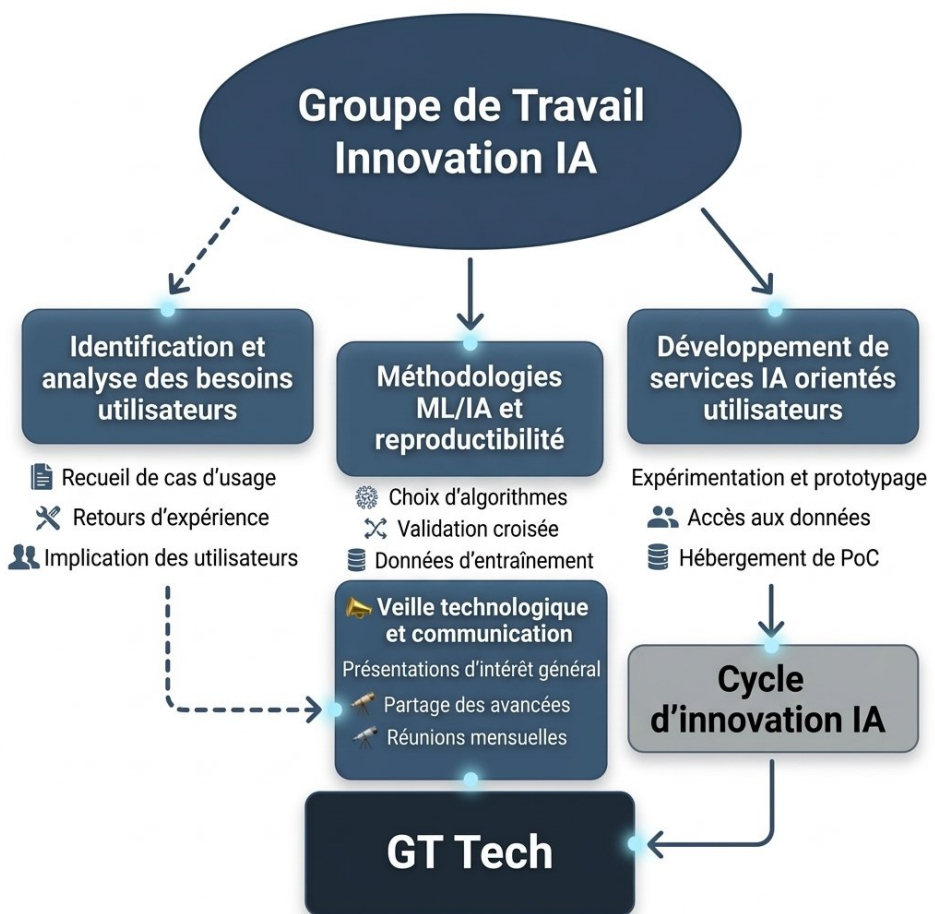


GT Innovation IA

- **Fédérer une communauté** de ~50 personnes autour des cas d'usage, approches et méthodes d'IA
 - ODATIS, CES, représentants PNDB, THEIA, AERIS, FormaTerre, Cellule Innovation Data Terra
 - Experts IA, représentants des utilisateurs finaux, responsables stratégiques



GT Innovation IA

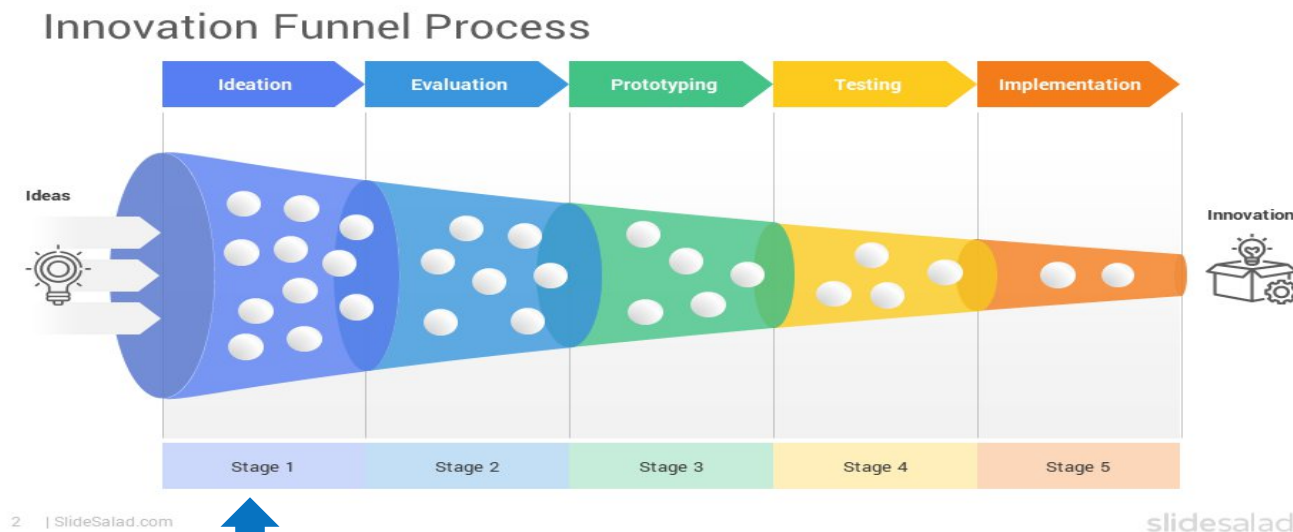


- **Fédérer une communauté** de ~50 personnes autour des cas d'usage, approches et méthodes d'IA
 - ODATIS, CES, représentants PNDB, THEIA, AERIS, FormaTerre, Cellule Innovation Data Terra
 - Experts IA, représentants des utilisateurs finaux, responsables stratégiques
- **Activités :**
 - Identifier et prioriser des cas d'usage
 - Instaurer un dialogue structuré et continu avec les communautés (scientifiques, techniques, stratégiques)
 - Échanger sur les actualités, les opportunités et les défis en IA
 - Etablir des liens avec les référents IA de l'IRD, Météo France, réseaux IA régionaux (ex ENV'IA)
 - Assurer le suivi de l'implémentation des cas d'usage prioritaires sur le plan fonctionnel et technique

Identifier et prioriser des cas d'usage

Besoin identifié	Transmis par	Description	Usages		Resources	Priorité 0..100
Annotation fiches catalogue	Clémence C.	Annoter et homogénéiser des fiches de catalogue. Premiers tests réalisés avec chatGPT	Interne			
Génération de rapports	Jacqueline B.	Genérer des rapports sur l'ensembles d'analyses d'un produit de données	Scientif			
AI Agent to perform actions	Cyril G.	Equiv de intercom.com/drlp/ai-agent - à mettre en place pour DataTerra	Interne +?			
Notebook LM	Sylvie F.	équivalent interne de Google NotebookLM : créer des chatbots spécialisés dans l'identification de poissons via des notebooks documentaires – un notebook par clé poisson	Scientif		<i>Open NotebookLM</i> comme alternative?	
Recherche sémantique	Thierry C. / Joël S.	Interrogation en langage naturel des SI scientifiques: ontologies, MCP pour exposer ses ressources aux LLM	Data users (externes ?)			

Identifier et prioriser des cas d'usage

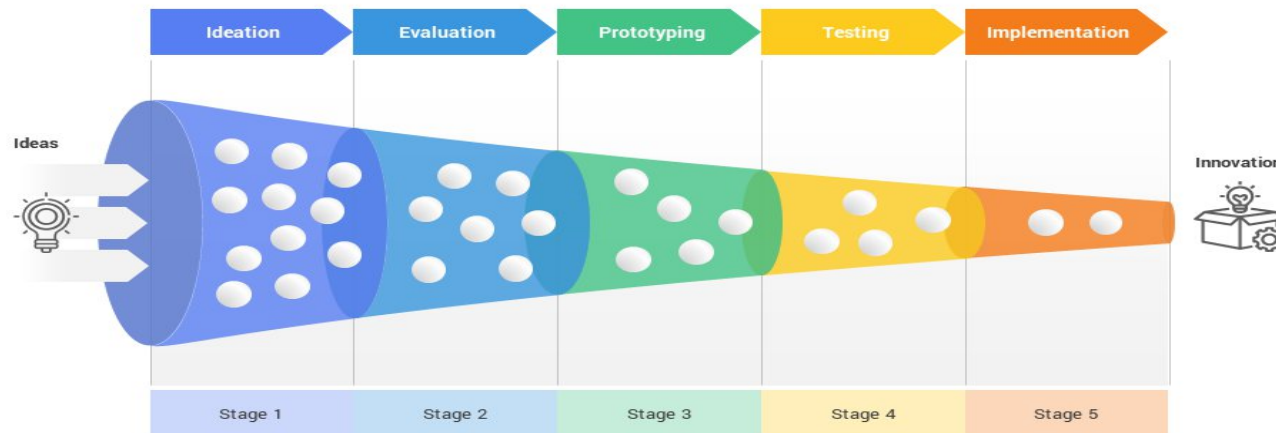


- IA dans la recherche
- Utilisateurs/Producteurs de données
- Décideurs publiques
- Partenariats privé

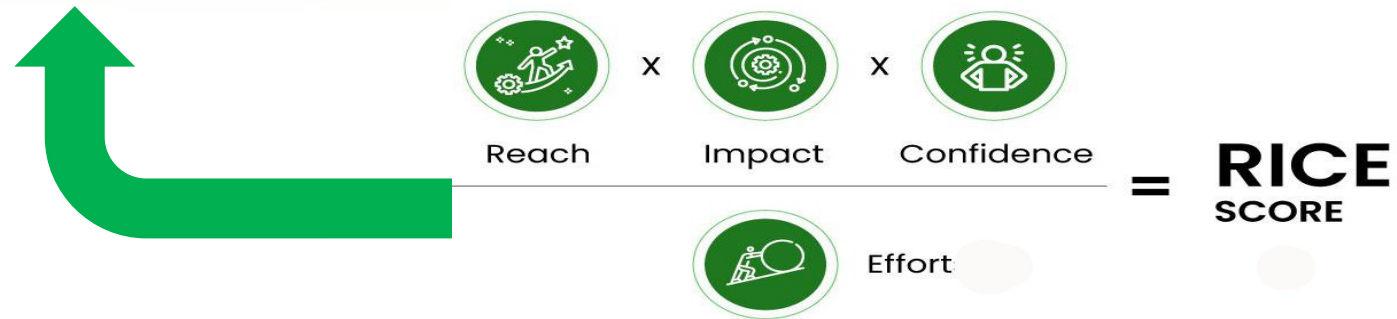
- ← Réseau référents labos
- GT IA ODATIS
- ← Ateliers ART,...
- ← Cellule Innovation DT

Identifier et prioriser des cas d'usage

Innovation Funnel Process



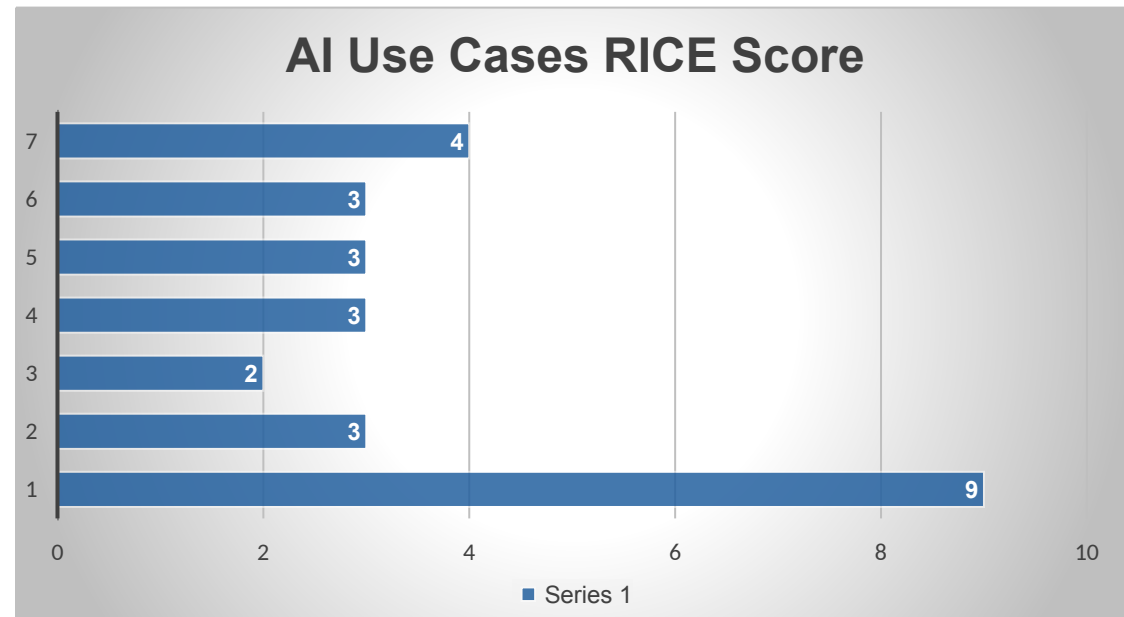
2 | SlideSalad.com



Identifier et prioriser des cas d'usage

Exemple d'exercice lors de l'atelier IA ODATIS, Jan 2026

- ✓ Le "Quick Win" : **Annotation de fiches** (Score 9). Impact direct sur la qualité des données et effort de développement modéré.
- ✓ Le "Long Terme" : **Agent IA** (Score 2). Complexité d'intégration et besoin en ressources élevé, abaissant le score RICE malgré un potentiel théorique fort.



Atelier IA janvier 2026 à MIO, Marseille

Vers la stratégie unifiée des Pôles

- Tour d'horizon des cas d'usage en Machine Learning et genIA au sein d'ODATIS et Data Terra
- L'état de l'art ainsi que les perspectives innovantes pour les organisations impliquées
- Identifier les axes stratégiques de travail et les priorités pour un développement coordonné et efficace



VERS LA STRATÉGIE IA D'ODATIS ET DES PÔLES DE L'IR DATA TERRA



Quatre thèmes d'océanographie appuyés par l'IA



DATA
TERRA



1. BIODIVERSITÉ & IMAGERIE ; SURVEILLANCE AUTOMATISÉE

- Identification taxonomique automatique d'organismes
- Analyse de tãraoctets d'images/vidéos benthiques *in situ*
- Utilisation par non-spécialistes (ex. EcoTaxa, DeepFaune)
- Suivi multi-source (images, sons, ADN, satellites)



2. PRÉDICTION & CLIMAT : ANTICIPATION DES RISQUES

- Prédiction d'événements extrêmes (cyclones tropicaux)
- Détection précoce d'efflorescences d'algues nuisibles (HABS)
- Emulateurs de modèles régionaux pour prévision météo-océanique
- Étude de variabilité du phytoplancton liée au climat (ex: CLEAR)

CLEAR



4. DONNÉES FAIR & LLM : ACCÈS ET DÉCOUVERTE DE DONNÉES

- Chatbots pour recherche documentaire (LLM / RAG)
- Enrichissement sémantique de catalogues (ex: ODATIS)
- Suggestion de mots-clés thésaurus, chatbots d'aide au déposant
- Développement de briques souveraines (ex: P16)



3. FUSION 4D & CONTRÔLE QUALITÉ : VISION CONTINUE DE L'OCÉAN

- Intégration BGC-Argo, satellites, modèles (4D - x,y,z,t)
- Cartographie des propriétés biogéochimiques (Chl-a, POC)
- Contrôle qualité automatique (Argo QC)
- Exploitation altimétrie haute



contact@odatis-ocean.fr | www.odatis-ocean.fr

Atelier IA ODATIS Janv. 2026

1. Biodiversité & Imagerie (Surveillance Automatisée)

- Reconnaissance automatique et la classification à haut débit des organismes marins (plancton et benthos)
- Remplace le travail fastidieux et coûteux d'identification manuelle par une analyse automatisée d'images massives (caméras sous-marines, microscopes, pièges photographiques)
- Permet de:
 - Étudier la biodiversité à des échelles spatiales et temporelles auparavant impossibles
 - Suivre la santé des écosystèmes en temps réel
 - Mieux comprendre les réseaux trophiques et le cycle du carbone (dont le plancton est un acteur majeur).

EcoTaxa aujourd'hui

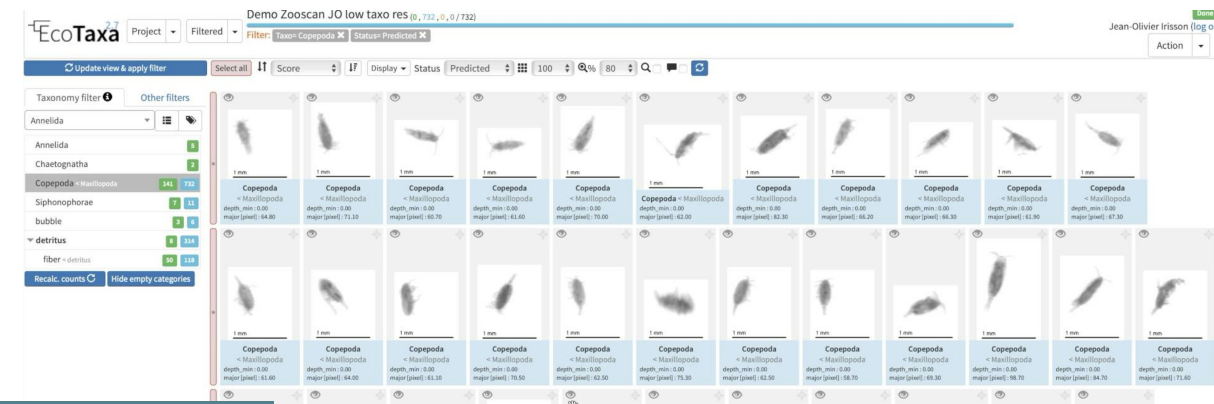
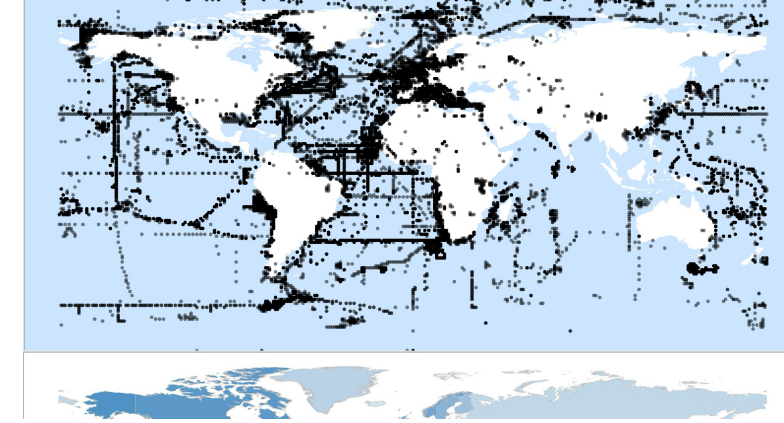
750M images, 250M validées par un opérateur humain

+15M par mois

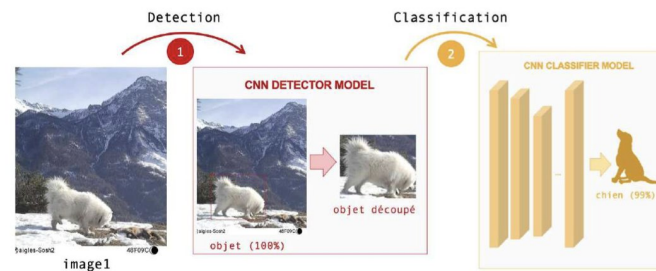
5,000 to 10,000 validations/opérateur/j

>4500 users from ~900 organisations

~125 utilisateurs concurrents



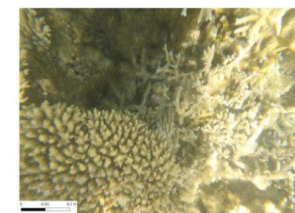
DeepFaune



Dimitry Khvoros

Habitat monitoring

« A transformer-based deep-learning model is trained on underwater images to detect the presence of 31 classes covering various coral morphotypes, associated fauna, and habitats »



Pl@ntnet: "Le Shazam des plantes"



DATA TERRA

ODATIS

2. Prédiction & Climat (Anticipation des Risques)

- Prédiction précise des événements extrêmes
 - L'IA permet d'intégrer et d'analyser en temps réel une multitude de données hétérogènes (satellites, bouées, modèles climatiques) pour identifier des schémas subtils que les modèles déterministes classiques peinent à saisir à court terme
 - Au-delà de l'enjeu scientifique de la compréhension du climat:
 - Améliorer les systèmes d'alerte précoce pour les populations côtières
 - Planifier les interventions d'urgence
 - Protéger les infrastructures
 - Limiter l'impact des algues toxiques sur les activités de pêche et de tourisme
- L'IA au service de la résilience des territoires

Reconstructing the Life Cycle of Tropical Cyclones

Etape 1 déterministe :

Apprend la relation ERA5 -> HWRF

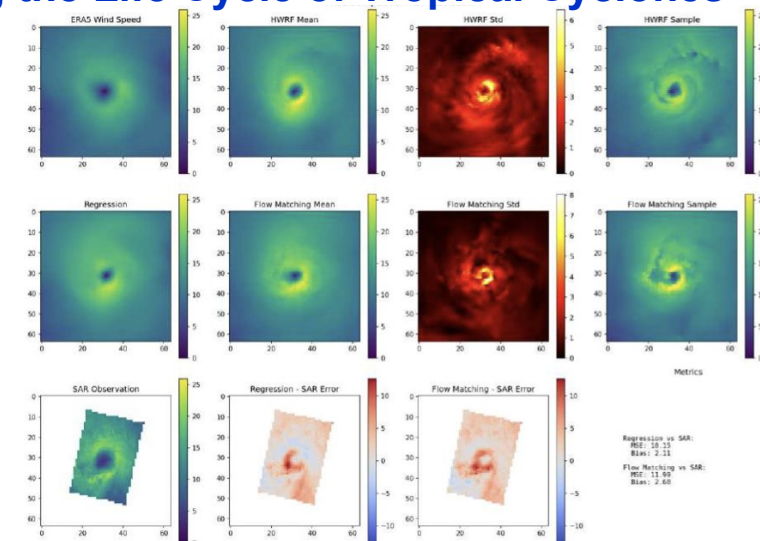
Rehausse les gradients de vents

Etape 2 probabiliste :

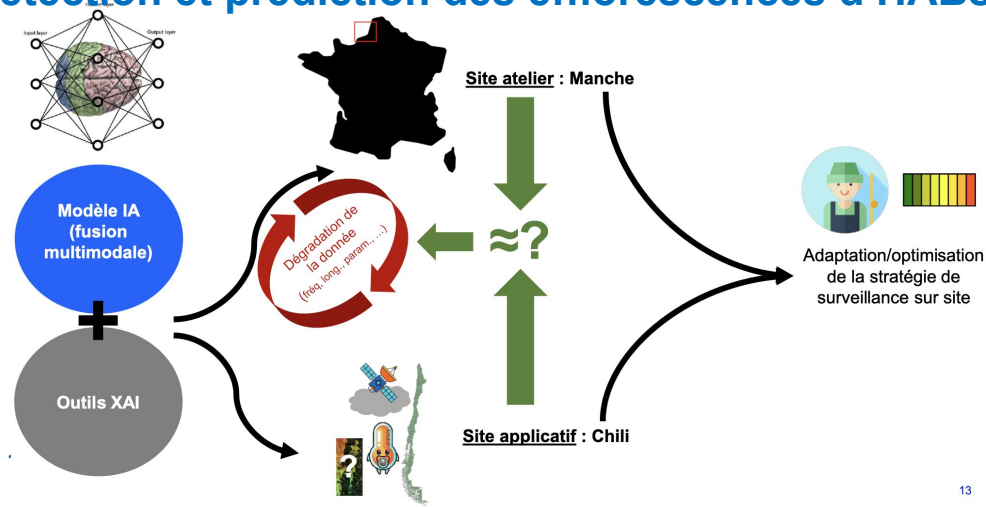
Utilise la sortie déterministe + bruit blanc

Apprend les résidus non expliqués

Génère une distribution des vents possibles



Détection et prédiction des efflorescences d'HABs



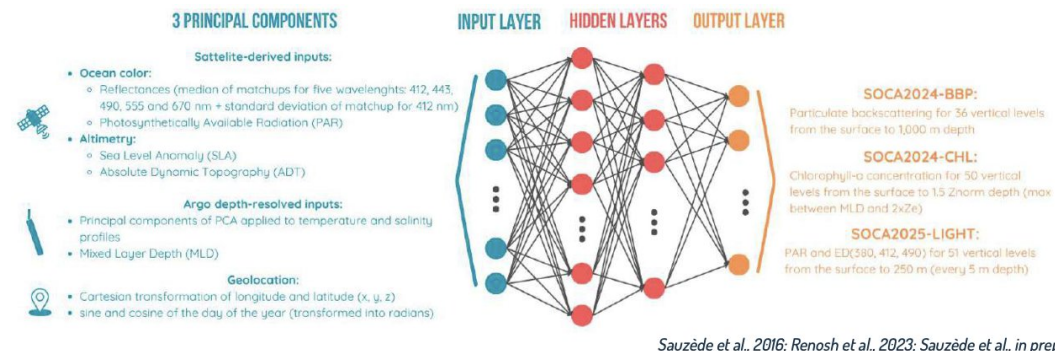
3. Fusion 4D & Contrôle Qualité (Vision Continue)

- Reconstruction 4D (espace-temps-profondeur) des propriétés biogéochimiques de l'océan, via la fusion de données satellites avec les données in situ (flotteurs Argo, programme BGC-Argo)

➤ Fin du problème du "pixel manquant" ? L'IA comme "ciment intelligent" : interpoler et extrapoler les données => créer des volumes continus (4D) de concentration en chlorophylle, oxygène, carbone,...

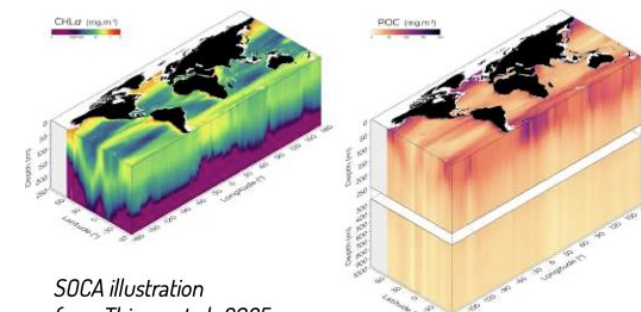
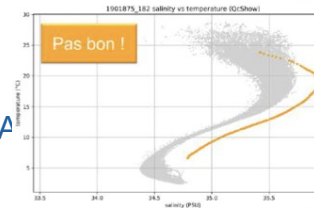
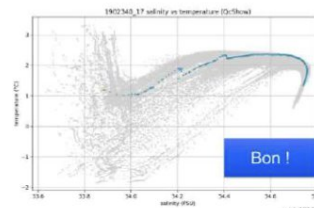
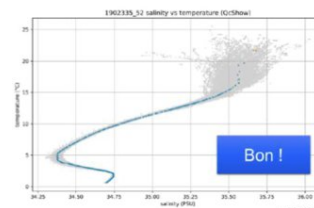
- ✓ "Tableau de bord" global, en profondeur et dans le temps, de la santé biogéochimique de l'océan.
- ✓ Aide à comprendre la pompe à carbone océanique (un enjeu climatique majeur)
- ✓ Suivre l'acidification des océans
- ✓ Modéliser l'avenir des écosystèmes marins

- Le contrôle qualité automatisé (Argo QC) garantit la fiabilité de cette vision 4D



Copernicus ID:

MULTIOBS GLO BIO BGC 3D REP 015 010

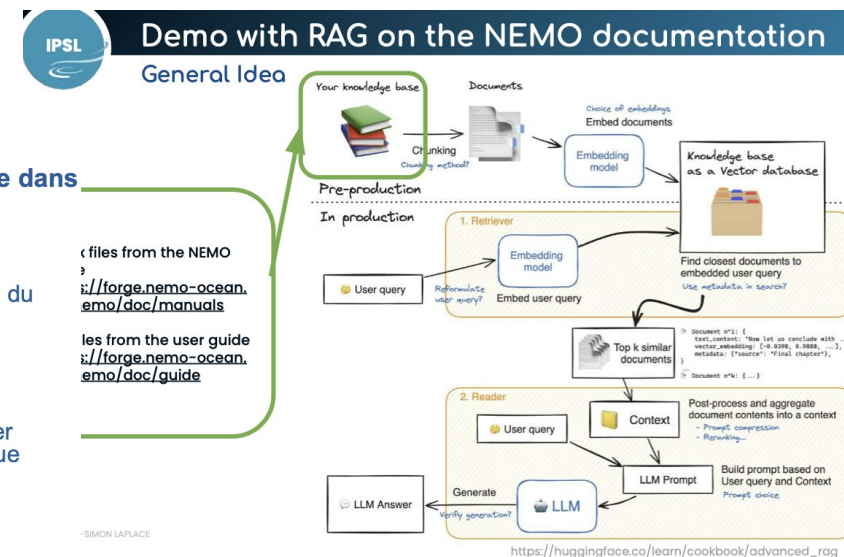


4. Données FAIR & LLM (Accès et Découverte)

- Intégration de Modèles de Langage (LLM / RAG - GenAI) pour l'aide à la découverte de données dans des catalogues massifs et complexes (ex. Sextant)
- ✓ Interroger les bases de données en langage naturel (chatbot). Le système comprend l'intention de la recherche, synthétise l'information issue de multiples sources et oriente vers les jeux de données les plus pertinents
- ✓ "Déverrouiller" l'accès aux montagnes de données existantes.
 - ✓ Un écologue peut trouver des données de physique sans être un expert du format NetCDF
 - ✓ Un décideur peut obtenir une synthèse rapide sur un enjeu local.
- Le concept de données FAIR devient une réalité pratique et intuitive !

Gestion des métadonnées : besoins pour l'amélioration de la découverte dans un catalogue

- 1^{ère} idée : aide au déposant
 - Suggérer des mots-clés des thésaurus Odatis au déposant à partir du contenu du titre et du résumé d'un jeu de données ou ressource (comme implémenté dans EaSy Data)
- 2^{ème} idée : aide à l'administration
 - Outil dans l'administration du catalogue permettant de demander à l'IA d'ajouter des mots-clés sur un lot de fiches de manière automatique ou semi-automatique (vérification humaine avant validation)
- 3^{ème} idée : vers une nouvelle manière de penser le catalogue
 - Facettes réduites à des informations « de base » (période temporelle, zone géographique, ...)
 - Recherche textuelle améliorée par l'IA et chatbot en complément du catalogue



Quelques tests effectués avec chatGPT : résultats probants

Nutrients and phytoplankton pigment concentrations at the Porticciolo station (Corsica Channel, Corsica)

DATE: 2025-05-12
TEMPORAL EXTENT: 2019-03-25 - 2024-11-28
AUTHORS: Goffart Anne¹
AFFILIATIONS: 1. Oceanology, FOCUS Research Unit, University of Liège, Belgium
DOI: 10.17882/106439
PUBLISHER: SEANDE

This dataset contains time series (2019 - 2024) of monthly nutrient and HPLC phytoplankton pigment concentrations measured at the Porticciolo station, a coastal station situated in the Corsica Channel (Northern Tyrrhenian Sea) over the 5 m isobath. The analyzed nutrients (in μM) are nitrate (NO_3^-), nitrite (NO_2^-), ammonia (NH_4^+) and silicate ($\text{Si}(\text{OH})_4$).

The measured pigments (in ng L^{-1}) include chlorophyll a (chl a), divinyl chlorophyll a (Dvchl a), total chlorophyll a (Tchl a), peridinin (Per), 19'-butanoyloxyfucoxanthin (Buta), fucoxanthin (Fuco), neoxanthin (Neo), prasinoxanthin (Prasino), violaxanthin (Viola), 19'-hexanoyloxyfucoxanthin (19HF), alloxanthin (Allo), zeaxanthin (Zea), lutein (Lutein) and total chlorophyll b (Tchl b). Sampling and analysis methods are described in Goffart et al. (2016).

These data contribute to characterize seasonal patterns of phytoplankton biomass and composition in a region where very limited information on phytoplankton dynamics is available. The data are also used to calibrate and validate ocean color remote sensing algorithms and to infer phytoplankton functional types using machine learning approaches.

Correspondance avec OD1

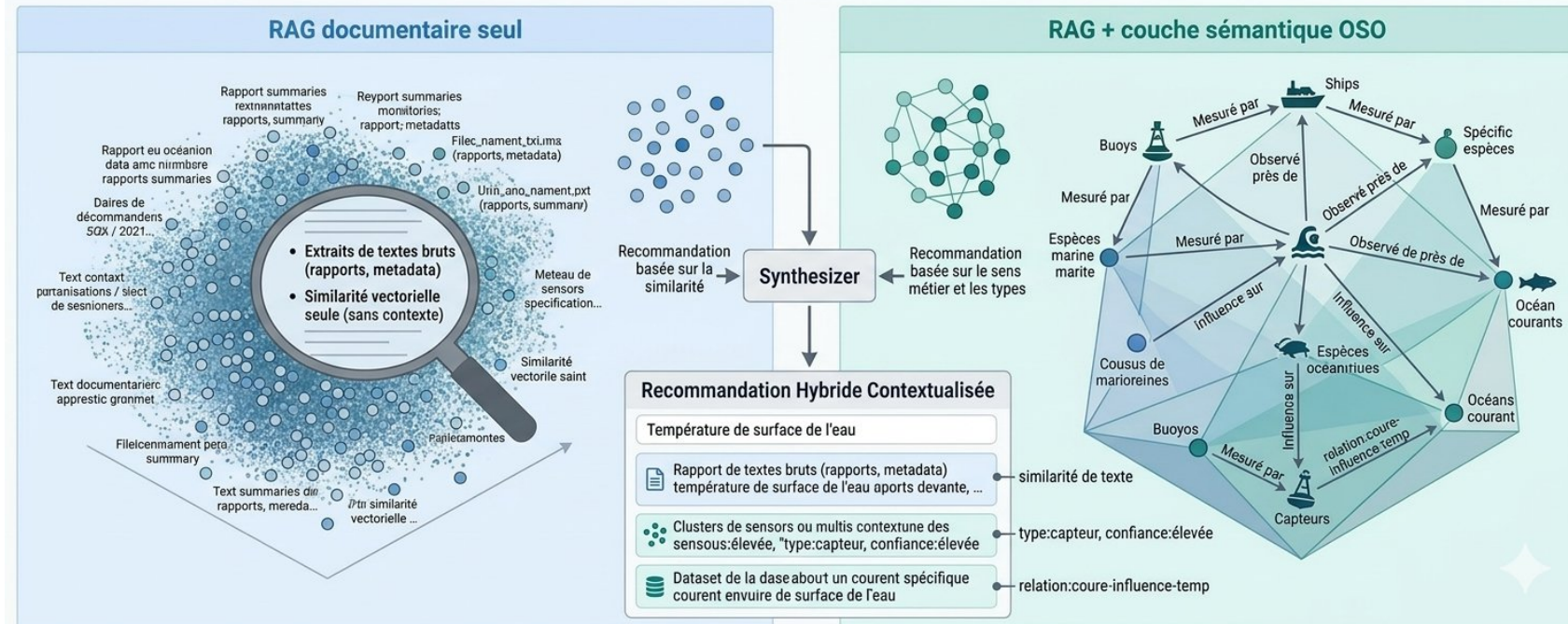
Code OD1	Libellé	Justification
NUTS	Nutrients	Englobe NO_3^- , NO_2^- , NH_4^+ , $\text{Si}(\text{OH})_4$ mesurés dans la colonne d'eau
PIGH	Pigments	Mesures HPLC de chlorophylles, caroténoïdes, autres pigments
PHYTO	Phytoplankton	Données sur la composition, biomasse, types fonctionnels

Dimitry Khvorostyanov - J

Cas d'usage : GraphRAG et moteur de recommandation

Architecture de Recommandation Hybride pour la Découverte Océanographique — Données "Beyond Text"

Cas d'usage commun



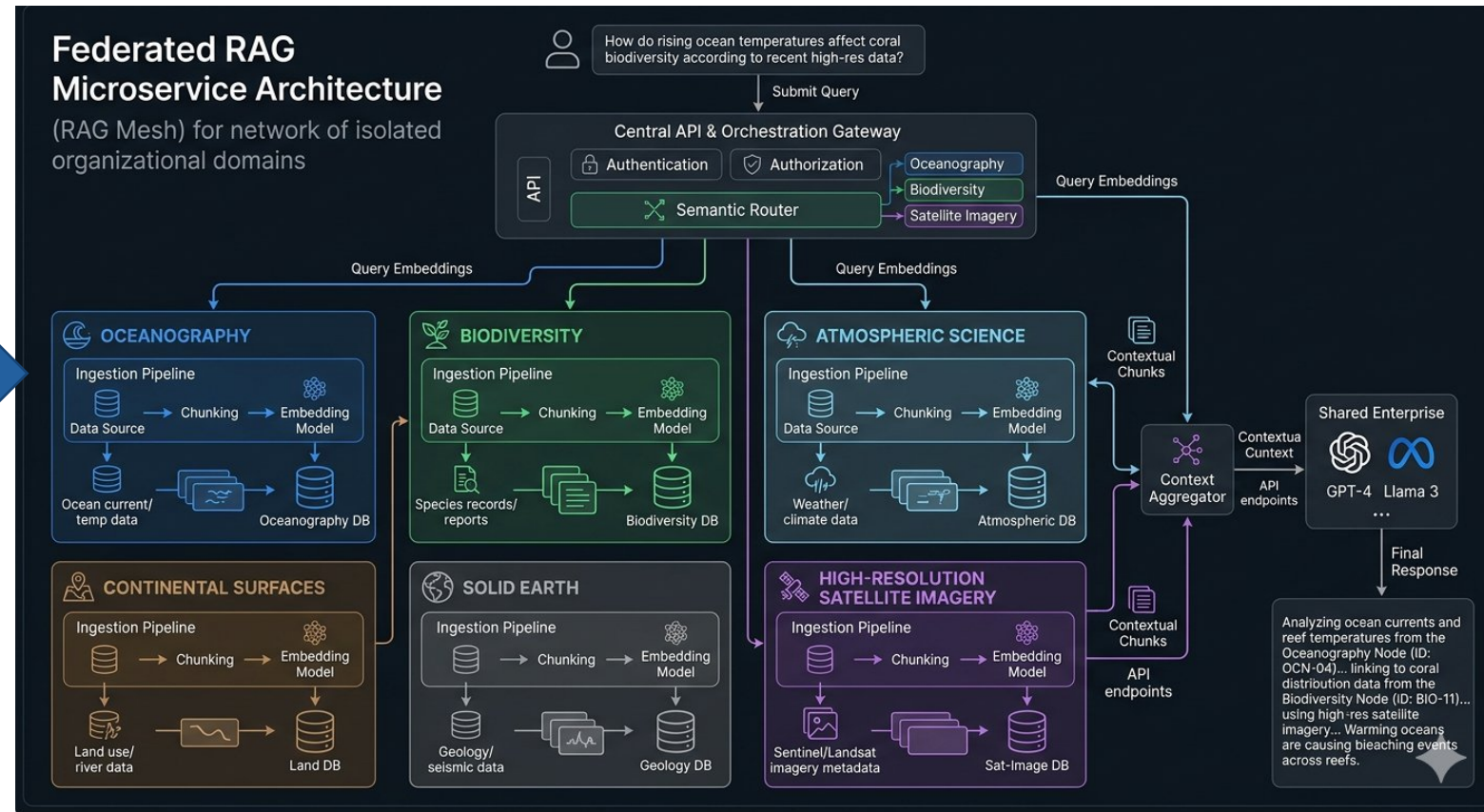
Stage Achille Djakpo (6 mois, ODATIS/LOCEAN), co-encadré Dimitry et Mylène

Collaboration étroite et contributions :

Francis Clément (PNDB), Charlene Guillaumot (THIEIA), Ibrahim Mohammad (Ifremer), Joël Sudre (Data Terra), Steven Piel (SISMER), Christian Pichot (INRAE), Christelle Pierkot (Earth Portal), Redouane Lguensat (AERIS), Fabien Gandon et Franck Michel (INRIA), Jean-Philippe Malet (FormaTerre), Jian-Sheng Sun (PNDB), Thierry Carval (ODATIS), ...

Cas d'usage : GraphRAG et moteur de recommandation

Vers l'architecture commune



Cas d'usage : GraphRAG et moteur de recommandation

Vers l'architecture commune



Sécurité Stricte

Cloisonnement étanche par domaine scientifique et isolation totale des données.



Indexation Sur-Mesure

Optimisation des embeddings selon la nature des données (géospatiale vs texte).



Performance Garantie

Élimination du "Noisy Neighbor" avec une scalabilité horizontale indépendante.



Gouvernance Locale

Chaque organisation reste maîtresse de son infrastructure et de ses coûts.



Valorisation IA

Chaque entité valorise son expertise en affichant son propre système RAG.



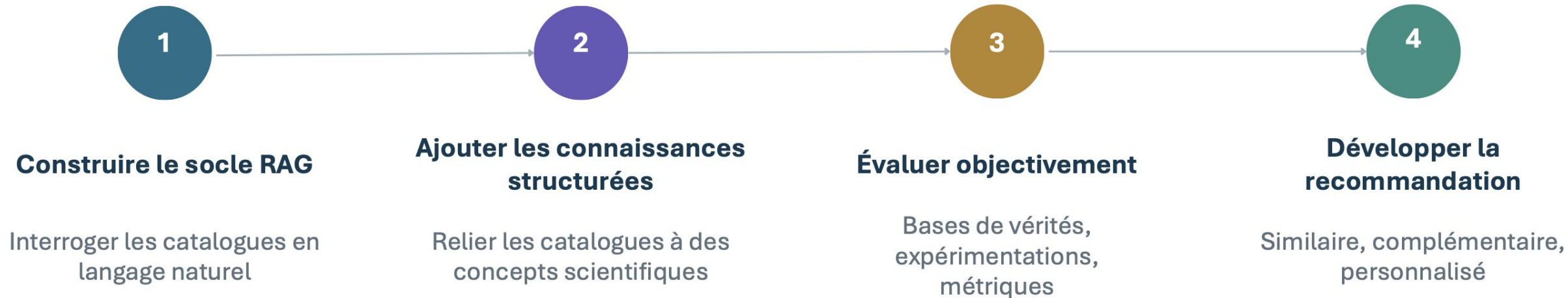
Intelligence Fédérée

Réponse aux requêtes complexes par orchestration sans fusionner les bases.

Cas d'usage : GraphRAG et moteur de recommandation

DÉMARCHE ET CONCEPTION

Trajectoire à suivre proposée



Fil conducteur : Partir d'un chatbot à une recommandation personnalisée

Cas d'usage : GraphRAG et moteur de recommandation

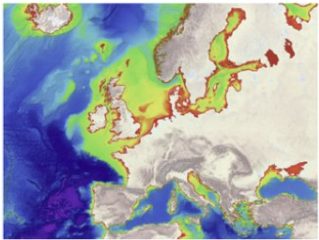
DÉMARCHE ET CONCEPTION

Analyse de l'existant

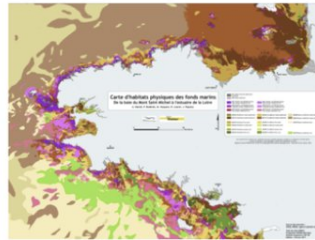
Les catalogues de données océanographiques

Sextant, Seanoe

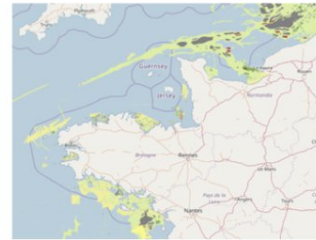
Découvrez les cartes interactives



Bathymétrie européenne



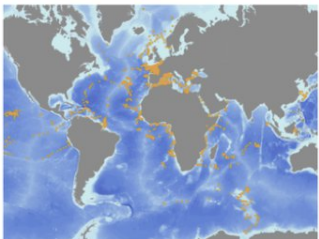
Habitats physiques



Granulats marins



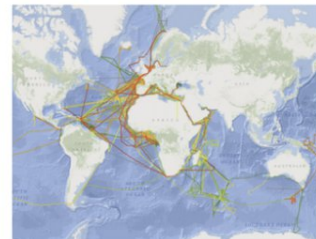
Hyperspectrale - La Réunion



Échantillons marins



Données des navires (géophysique)



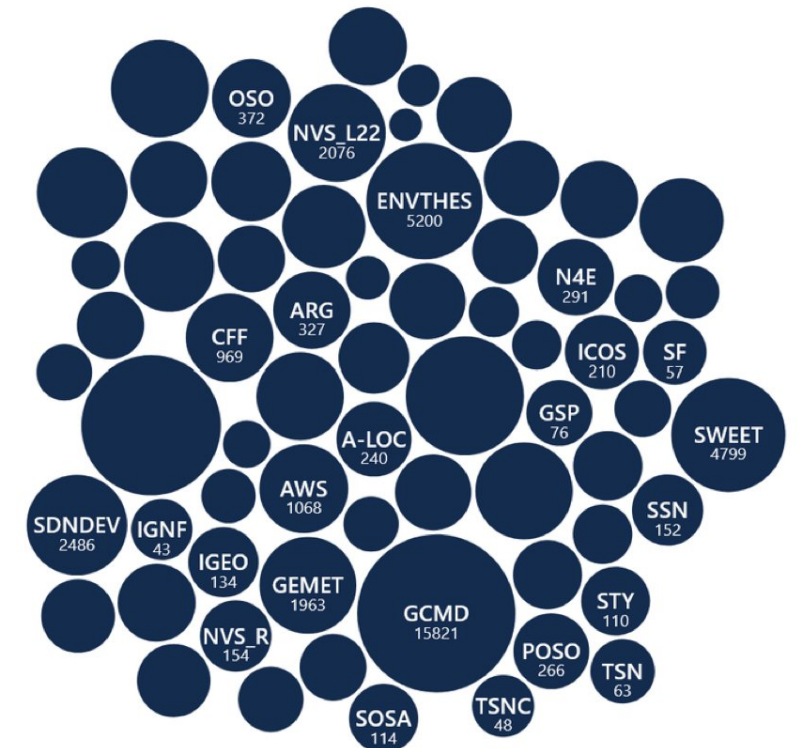
ADCP des navires (courants)



Typologie des habitats - Nouvelle-Calédonie

Les artefacts sémantiques

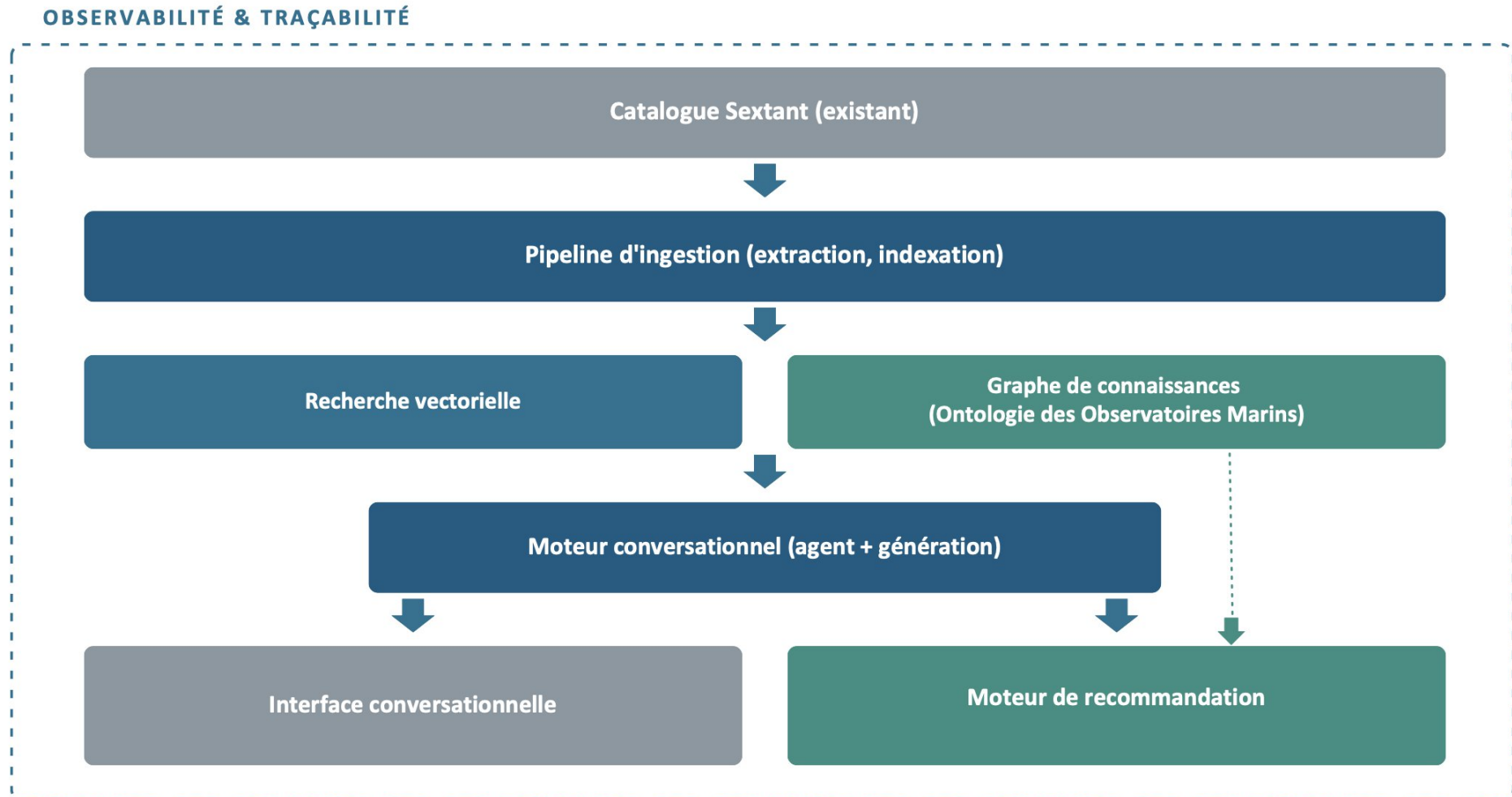
EarthPortal : Ontologies, Thésaurus...



Cas d'usage : GraphRAG et moteur de recommandation

DÉMARCHE ET CONCEPTION

Vue d'ensemble de l'architecture cible



Cas d'usage : GraphRAG et moteur de recommandation

A

Ariane

Assistant documentaire

+ Nouveau chat

HISTORIQUE

combien de fiches de données...
24/08/2026

https://vast-slim-tide-default-data-terra.ai.meso.umontpellier.fr

☆ | D Terminer la mise à jour

≡

Portail de découverte documentaire · Sextant

Explorez les catalogues de données marines

quelles sont les jeux de données de salinité SMOS

U

AI

Plusieurs jeux de données concernant la salinité de surface de l'océan issus de la mission **SMOS** (Soil Moisture and Ocean Salinity) sont disponibles, principalement via le centre de traitement **CATDS** (Ifremer).

Voici les principaux produits identifiés :

1. Produits de Salinité de Niveau 4 (Analyses globales)

Ces produits utilisent des algorithmes d'interpolation optimale (OI) pour fournir des champs de salinité continus en combinant les données satellites et les mesures *in situ*.

- **CATDS-PDC L4OS SMOS/SMAP-OI** : Produit hebdomadaire combinant les données des satellites SMOS et SMAP ainsi que le système d'analyse *in situ* (ISAS).
 - **Accès** : Des versions combinant SMOS uniquement (depuis 2010) et SMOS+SMAP (depuis 2015) sont disponibles via le portail [CATDS](#).
- **Produits de salinité de niveau 4 (L4)** : Le catalogue mentionne des analyses globales de la salinité de surface (SSS), de la densité de surface (SSD) et de la "spiciness" (SSSp).

2. Produits de Salinité de Niveau 3 (Champs moyennés)

Posez une question sur les jeux de données, organismes, zones, espèces... >

Projets récents à l'IR Data Terra et infrastructures

- GenAI4Earth
- AI2F - AI Factories - EuroHPC

Infrastructures Européennes

- **NFDI4Earth** (Gestion de données métier). Consortium allemand dédié aux géosciences et à l'environnement (réseau NFDI), moteur au niveau européen. Il fournit des données thématiques structurées selon les principes FAIR et gère les outils métiers propres aux sciences de la Terre.
- **AI4EOSC** (Plateforme de services IA / Middleware). Projet européen ancré dans l'EOSC qui développe des solutions logiciels pour l'IA. Il fournit aux chercheurs un environnement complet (MLOps) pour créer, entraîner et déployer des modèles de Deep Learning sur des infrastructures distribuées..
- **AI Factories** (Infrastructure & Puissance de calcul). Initiative européenne majeure s'appuyant sur le réseau de supercalculateurs EuroHPC. Elle fournit l'infrastructure matérielle lourde (clusters GPU, stockage très haut débit) pour entraîner des modèles d'IA très volumineux.
- **CDIF4EOSC** (Interopérabilité interdisciplinaire). Projet européen financé par Horizon Europe et coordonné par CODATA. Il fournit le cadre sémantique et les standards d'interopérabilité nécessaires pour croiser des données issues de disciplines différentes et les rendre exploitables par l'IA (FAIR-by-design).

GenAI4Earth : services d'IA générative et agentique pour le système Terre

Projet co-construit :  EOSC Node | Data Terra Environment

 EOSC Node |  nfdi
Nationale
Forschungsdaten
Infrastruktur

Coordination: CNRS / DATA
TERRA

17 partenaires / 8 pays: EOSC Nodes,
ESFRI-RIs, centres EuroHPC, SMEs &
universités

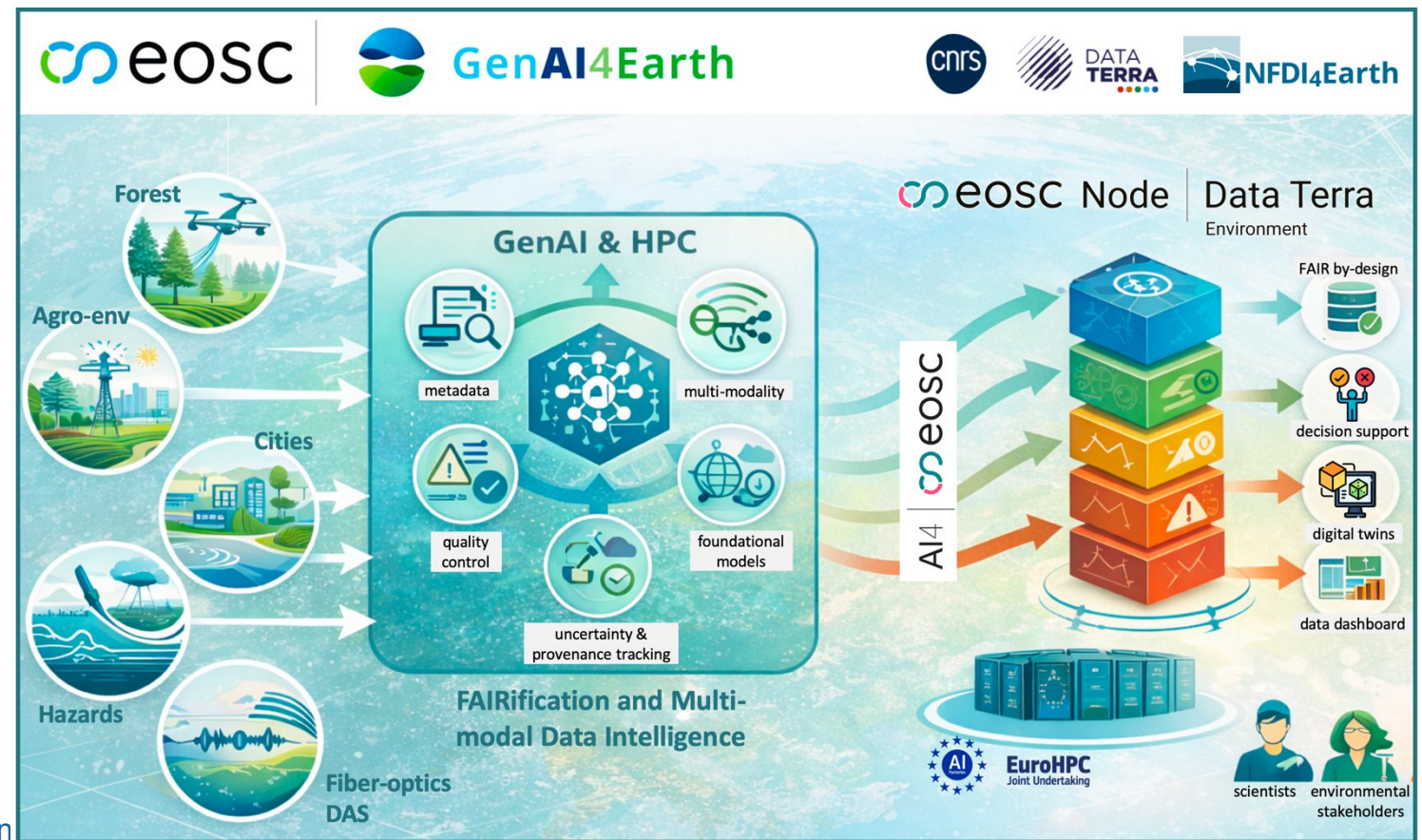
Support: AI Factories (FR-AI2F, DE-
HammerAI, PL-PIAST-AI), ERICs du
système Terre (EPOS, ICOS, eLTER,
AnaEE, Emphasis, EMSO)

Alignement: EOSC, ESFRI, GenAI4EU, AI
Act, digital sovereignty agenda

Budget: €9,976,151.18

PO: Anna Santoro

Dimitry Khvorostyan

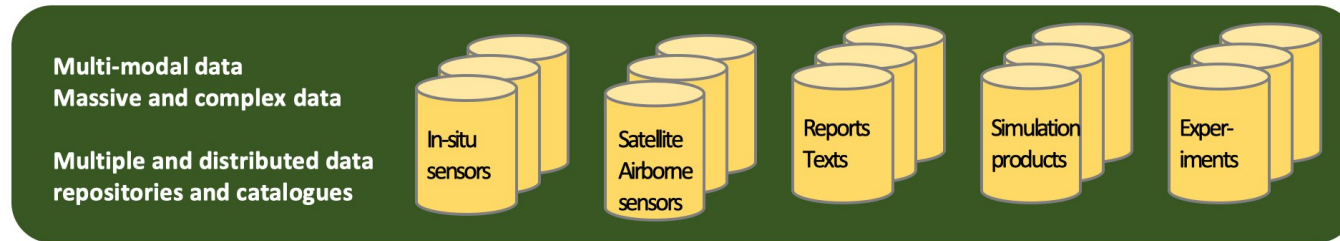


© designed with AI (Mistral)

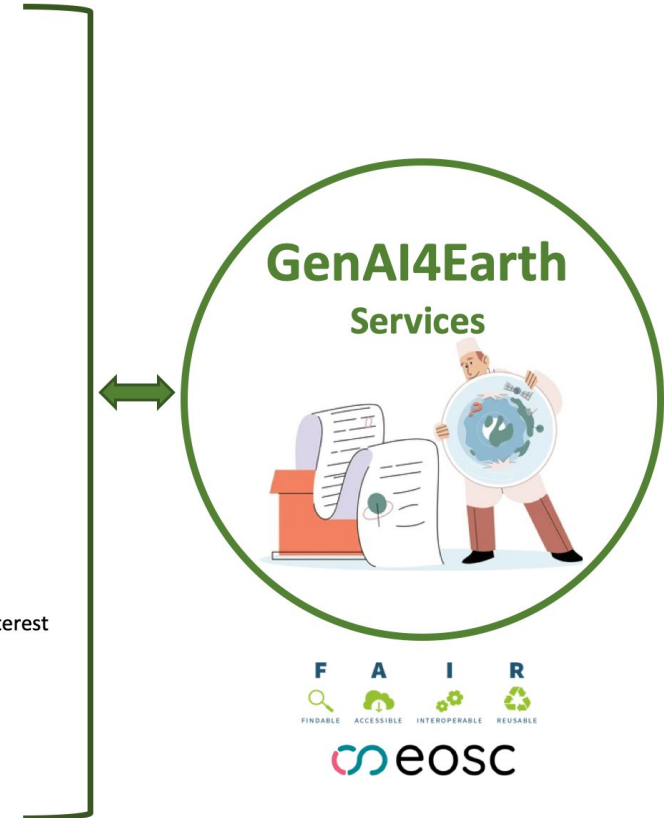
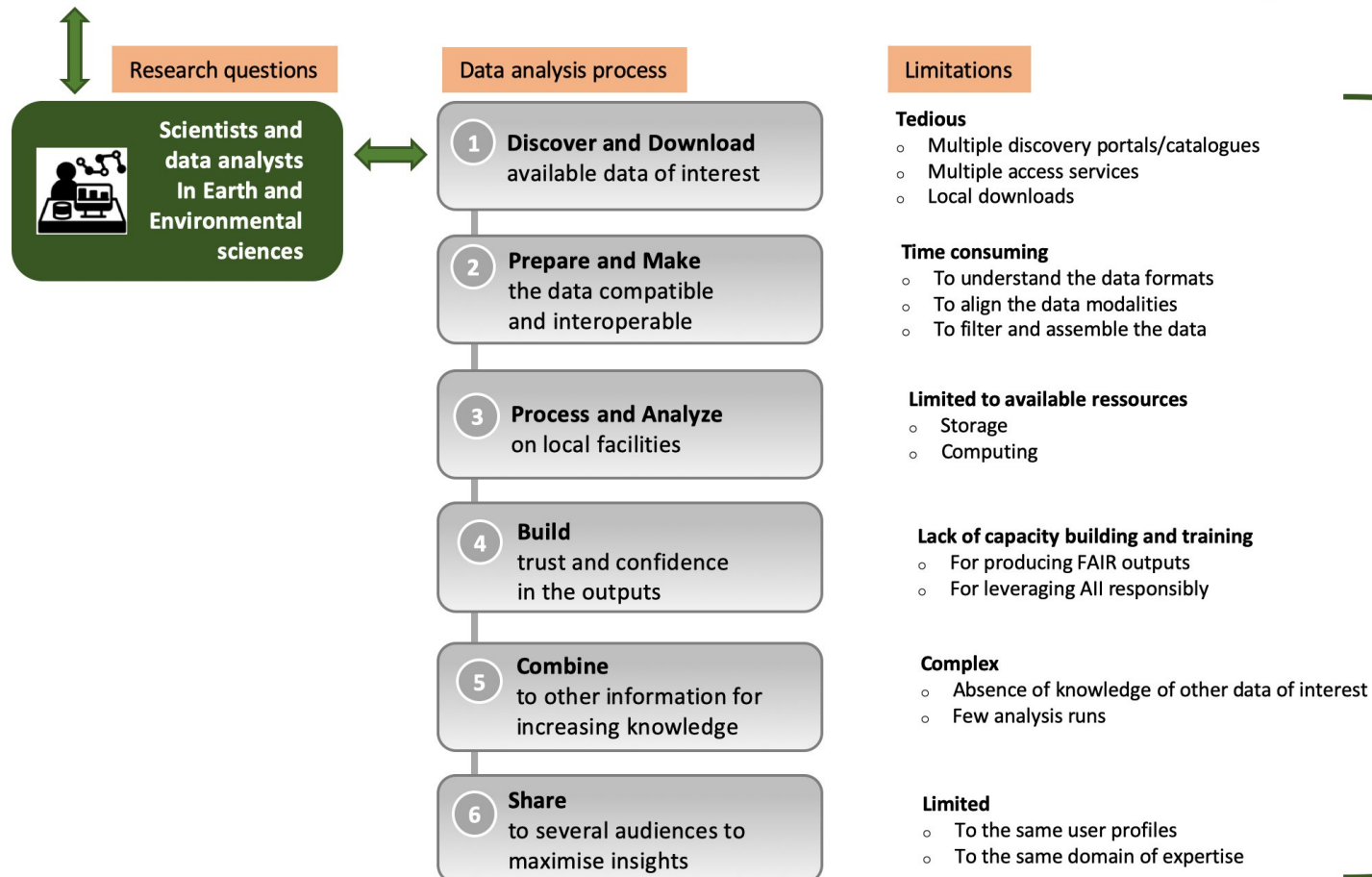
DATA
TERRA

ODATIS

Objectifs GenAI4Earth



Develop and implement specific GenAI workflows and AI agents to support Earth System Science



Objectifs GenAI4Earth

- Intégrer des services d'IA générative de confiance pour les communautés système Terre au sein de la Fédération EOSC : alignement avec les besoins exprimés Data Terra et NFDI4Earth
- Contribuer à l'écosystème AI4EOSC: 'AI4EOSC for ESS'
- **Services IAGen pour documenter les données « système Terre (Tâches 7.1/8.1)**
 - → Des outils et services IA pour augmenter le FAIRness, l'exploitabilité machine, l'aptitude à être exploité par l'IA des données
 - → nouveaux standards de description (Croissant/GeoCroissant), LLMs d'auto-complétion de métadonnées
- **Services IAGen pour la découverte des données « système Terre » (Tasks 7.2/8.2)**
 - → Développer et proposer une suite d'agents IA DATA TERRA entraînés sur des métadonnées de données du système Terre pour leur découverte, choix, accès, et pré-intégration dans des chaînes d'analyse et/ou environnements de recherche
 - → interopérabilité CDIF4EOSC ; liquidité – FLUID-AI
- **Services IAGen pour l'analyse croisée de données à travers des workflows scientifiques**
 - → « Fine-tuning » de modèles fondation (GeoFM/TimeFM) avec de nouvelles modalités (1. résilience urbaine / îlots de chaleur/pollutions atmosphériques, 2. surveillance agro-forestière – biomasse/carbone, 3. sismologie par fibre optique), etc.
- **Formations multi-profils utilisateurs**
 - → Ecole d'été DATA TERRA/NFDI4Earth tous les ans sur les méthodes d'IA Générative

GENCI : Infrastructure pour l'IA au niveau français

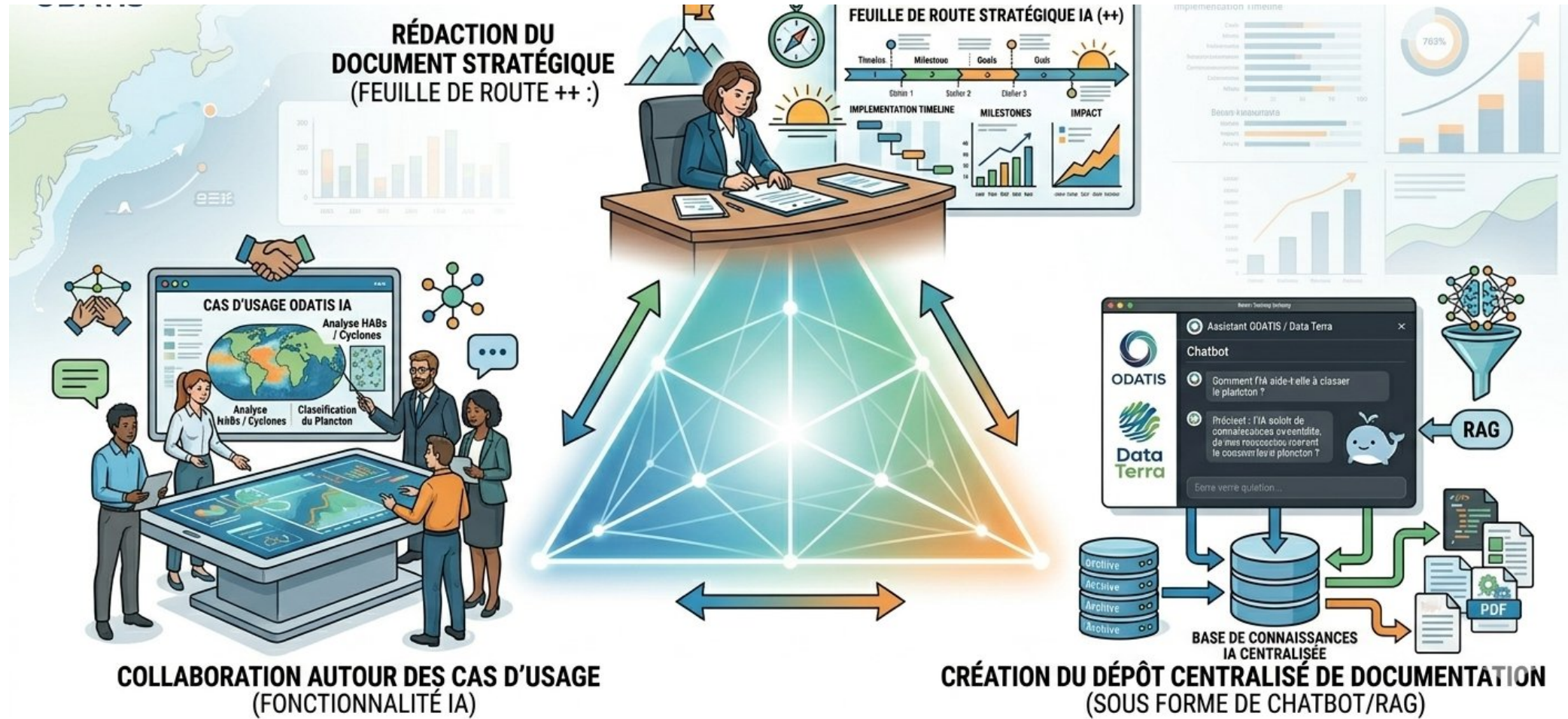
CINES & Supercalculateur Adastra

- L'un des 3 centres nationaux de calcul de GENCI (avec l'IDRIS et le TGCC).
- Héberge et exploite Adastra, l'un des supercalculateurs souverains les plus puissants pour l'IA et le HPC.
- Partenaire stratégique de Data Terra / GAIA Data (EquipEx+).

DrOCC-Est / ISDM-MESO

- Site partenaire de Data Terra / GAIA Data dédié à l'hébergement et au déploiement de services avancés.
- Plateforme d'expérimentation pour les technologies IA (ex. déploiement d'architectures RAG).

Prospectives



Prospective: Collaboration autour de cas d'usage

Exemple: Identifier les espèces marines : du besoin terrain à la solution collaborative

Expertise taxonomique (Dominique Ponton) : Guidage textuel précis via des critères méristiques et morphologiques, avec des notebooks spécialisés par région/famille.

Expertise algorithmique (David Mouillot) : Reconnaissance d'images (YOLO/CNN) et approche multimodale (texte + géolocalisation)

Vers un Hackathon ODATIS / PNDB

Fiabilité : Texte (guides experts) + IA (images) pour une identification fiable, même pour espèces abîmés

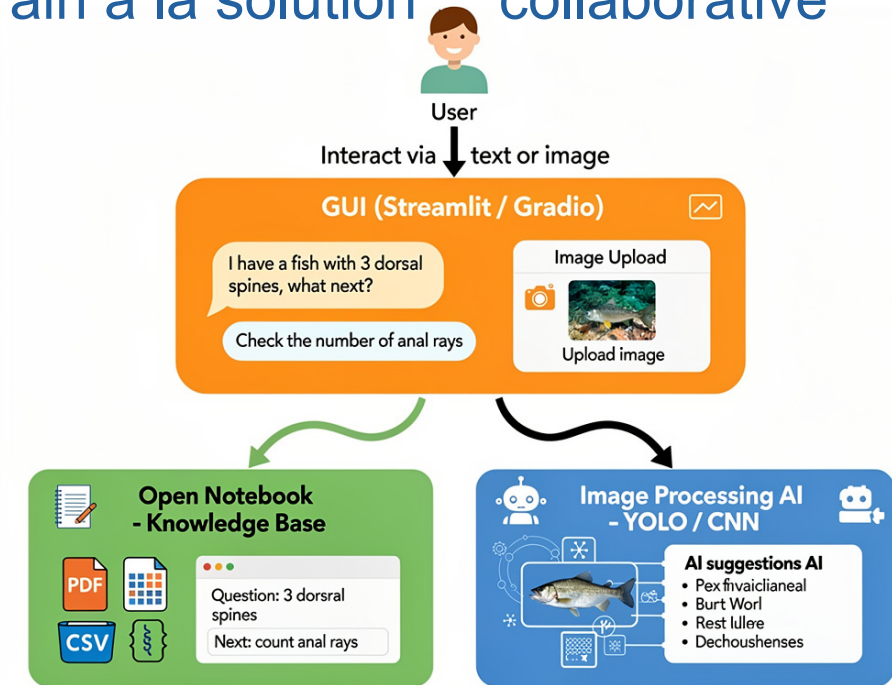
Simplicité : Interface intuitive pour non-experts (questions/ photos)

Modularité : Notebooks ciblés par région/famille pour éviter la surcharge

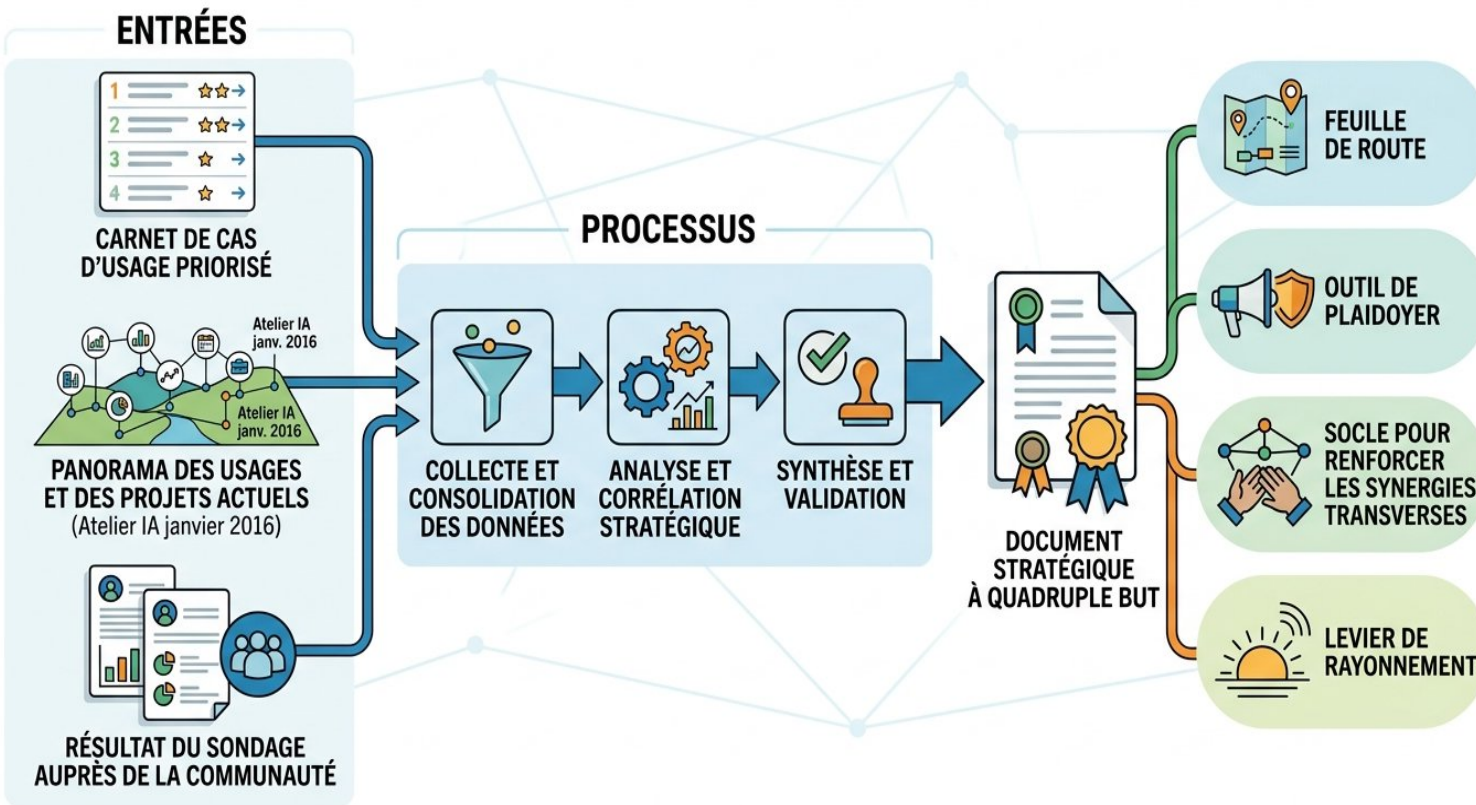
Collaboration : L'IA propose, l'expert valide et affine

Dimitry Khvorostyanov - Journée ODATIS Océan Indien – 18/09/2026

| 30



Prospective: rédaction du document stratégique IA



- *Feuille de route pour piloter le développement de l'offre IA*
- *Argumentaire stratégique pour mobiliser les ressources auprès des tutelles, appels d'offre,...*
- *Socle pour renforcer les synergies transverses entre les pôles*
- *Levier de rayonnement pour positionner la communauté au niveau national et international*

Prospective: centraliser la documentation

Ressources, bonnes pratiques et compétences en IA

Vers la documentation interactive centralisée

Ressources
IA

temperature.

oceanography.
salinity, currents

depth, depth

Quelle est l'occurrence
d'utilisation des algorithmes IA
dans des pôles de Data
Terra ?



Prospective: centraliser la documentation

Ressources, bonnes pratiques et compétences en IA

Vers la documentation interactive centralisée

Ressources
IA



Quelle est l'occurrence d'utilisation des algorithmes IA dans des pôles de Data Terra ?

