

Retour des ateliers des 10 ans AERIS

Atelier 1 : Les données spatiales au service de la communauté atmosphère (Hervé Roquet, Carole Deniel)

On constate une augmentation de différentes plateformes web d'accès aux données spatiales opérées par des agences spatiales européennes (ex EWC, WEkEO) ainsi qu'une augmentation attendue du volume des données spatiales liée à une forte hausse des missions spatiales à venir. Ces plateformes web peuvent proposer des ressources de calcul ainsi que des services (mais pas encore aussi performants que ceux d'AERIS). Cet environnement de données spatiale est très complexe pour les utilisateurs scientifiques et l'un des intérêts d'AERIS serait de **faciliter l'accès à ces plateformes et à accompagner leur utilisation par la communauté scientifique** (par exemple : aide à l'utilisation des environnements virtuels et des ressources de calcul). Il est important pour AERIS de ne pas dupliquer les données spatiales : ainsi pour la récupération des données IASI-NG, AERIS s'appuiera sur l'EUMETCAST terrestre disponible au Centre de Météorologie Spatiale à Lannion.



Il y a une demande d'**assistance des utilisateurs scientifiques français pour la soumission de projets aux plateformes européennes**. AERIS peut centraliser les demandes de la communauté française et aider le CNES/Météo-France à valoriser les demandes de contribution nationales tout en veillant à ce que le processus ne soit pas lourd pour les utilisateurs.

Il y a eu une présentation de la plateforme GEODES du CNES. Cette plateforme, nœud de GAIA DATA présente un intérêt pour la communauté Data Terra mais peu pour les utilisateurs d'AERIS (sauf pour le croisement de données avec d'autres thématiques).

Il y a eu durant la discussion une question sur la **formation des scientifiques au calcul HPC** ou cloud (le CNES finance des contractuels dans plusieurs laboratoires dans le cadre de la mission SWOT). L'expertise sur le calcul HPC devrait revenir à AERIS et le pôle doit être intégré en amont des montages de missions spatiales avec les laboratoires pour veiller au mieux à ce que les chaînes de traitement développées par les scientifiques soient le plus adaptées possibles aux ressources d'AERIS.

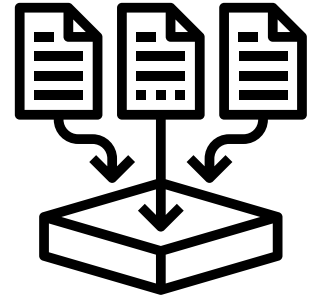
Enfin, concernant la question sur **l'offre de données d'agences étrangères**, AERIS doit préciser les mécanismes qui permettent aux scientifiques d'avoir un support pour l'accès aux données d'agences étrangères (Chine, Inde, Japon...) voire aux données du domaine privé. Il est préférable de passer par l'appel à projet AERIS pour ces demandes là (AERIS ne va pas distribuer spontanément des données d'agences étrangères s'il n'y a pas une demande scientifique).

Atelier 2 : L'évolution des outils et services AERIS

(Guillaume Brissebrat, Fabrice Ducos, Sophie Cloché, Damien Boulanger)

L'objectif de cet atelier est de discuter sur les retours des utilisateurs d'AERIS sur les différents outils et services existants et de discuter des évolutions en cours et à venir (prospective sur les 5 prochaines années). Les services aux données d'AERIS comprennent :

- Découverte (portails web et catalogues)
- Accès (téléchargement "basique", services de données, outils plus avancés de type colocalisation/extraction...)
- Visualisation (Dataviz, support aux campagnes...)
- Calcul (cluster utilisateur AERIS/ICARE...)



Il y a une demande d'**outils de visualisation et d'extraction des données de réanalyse ERA5** pour la communauté atmosphère. Ces données sont utiles à l'ensemble de la communauté Data Terra, une réflexion est en cours pour traiter cela dans le cadre du projet Gaia Data. Il y a également une forte demande de **prévisualisation des données dans le catalogue AERIS**. Cette action est en cours d'implémentation dans le catalogue AERIS, mais il y a la nécessité de standardisation des données en entrée. Le catalogue devra aussi renvoyer vers des outils plus avancés quand ils existent (Dataviz ICARE par exemple)..

Une demande concerne la **création d'outils de détection d'anomalie dans les séries temporelle**. Des outils (REOBS...) sont déjà en cours de développement dans le cadre d'ACTRIS pour le contrôle qualité des séries longues ou la détection de trous dans les données transmises mais ils ne sont pas génériques à l'ensemble des données AERIS.

Une demande formée par les utilisateurs d'AERIS est celui d'un **outil générique pour le contrôle qualité des données**. Un outil est actuellement en cours de développement pour les producteurs de données d'ACTRIS-FR mais il ne sera pas automatique, la saisie sera manuelle. Des processus sont mis en place par AERIS pour détecter les incohérences dans les données (mais cela devrait être du ressort du producteur de données)



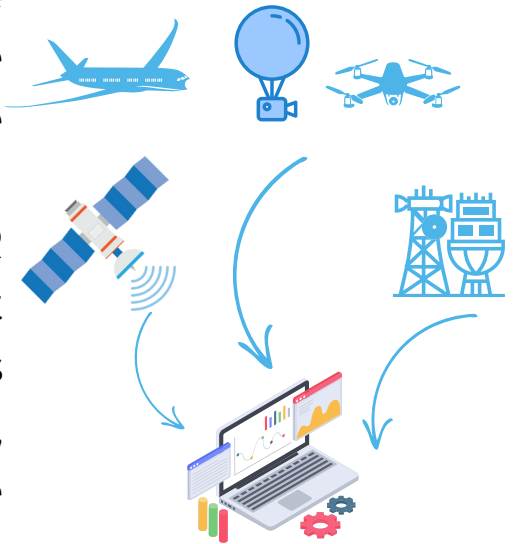
Aperçu de l'outil de visualisation Station Hub pour les données des plateformes basées au sol (ici un exemple avec une station ACTRIS ARES)

Enfin, une attention a été portée sur la **communication au grand public/médias** : il y a un besoin de **jeux de données de références AERIS** pour répondre aux médias sur des événements extrêmes. Ainsi que faire des outils de visualisation simple (en s'inspirant de Météo-France) pour la communication grand public. Il est important de faire découvrir les outils/services d'AERIS par les ateliers de la donnée et école thématiques atmosphère.

On peut noter également un besoin de **dialogue entre la communauté modélisatrice et observation** (par exemple en facilitant l'accès, l'utilisation et la documentation de jeux de données d'observation pour les modèles, en adoptant des formats type NETCDF, en construisant des jeux de données plus faciles à utiliser pour des communautés non spécialistes des observations (comme les données FROGs, GEWEX-CA, REOBS par exemple)

Atelier 3 : L'importance croissante du croisement des données : quels besoins, quels outils ? (Éric Defer, Martial Haeffelin, Cyril Crevoisier)

Il existe déjà de nombreux **produits chez AERIS résultant de croisement entre différents capteurs satellitaires** (comme par exemple les profils verticaux d'ozone provenant de données IASI et GOME2 ou encore les produits SODA et DARDAR provenant de données CALIPSO/Lidar et CLOUDSAT/radar), mais également des **produits de données de réseaux sol**, comme le produit GAARLIC sur la propriété des aérosols qui mêle des données de lidars et photomètres.



Il existe également des **services aux données pour faciliter leur croisement** comme par exemple l'outil de colocalisation/extraction de données de produits satellite avec des données réseaux sol ou encore des outils de projection de données sur une même grille spatio-temporelle facilitant leur comparaison (par exemple le produit FROGS pour l'ensemble de produits de précipitations) ainsi que des **plateformes web thématiques comme Volcplume** qui permet la visualisation et manipulation de données diverses (IASI, Sentinel 5P, imageurs géostationnaires, réseaux sol...) pour le suivi des panaches volcaniques.

Pour faciliter le croisement de données, plusieurs demandes ont été faites durant la table ronde :

Il y a une demande de **visualisation des données disponibles dans AERIS** pour voir si elles sont intéressantes pour un croisement, pour faire émerger des questions scientifiques et pour la détection d'anomalie. AERIS doit donner **accès à des environnements pour faire de la visualisation spécifique par produits** (chez AERIS/ICARE il est possible de précalculer des visualisations et de faire de la visualisation à la demande).

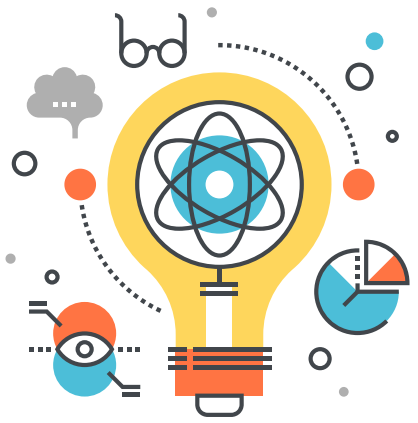
Il y a également un besoin d'**aide au croisement des données entre les différents pôles de Data Terra** : AERIS doit fournir des outils pour la récupération des données entre les différents catalogues des pôles. Un besoin de coordination est apparu sur les vocabulaires et les formats entre les différents pôles de Data Terra.

Une demande de **partage d'expérience entre les scientifiques** sur le croisement de données est demandée (par exemple via la création d'une ressource) ainsi que sur le **partage d'outils spécifiques mis à disposition d'AERIS** (par exemple ce qui est fait dans le cadre de la mission spatiale C3iel ou bien les outils développés dans le cadre du support aux campagnes) ainsi qu'un espace de dépôt de code opérationnel pour les scientifiques.

Enfin, un besoin a été demandé d'assistance par AERIS sur le **retraitement de données anciennes** (il y a une expérience dans le centre AERIS/ICARE grâce au projet de Copernicus C3S Data Rescue).

Atelier 4 : De la production à la valorisation des données de la recherche (Stéphane Sauvage, Jérôme Riedi)

L'objectif de l'atelier était de voir comment les infrastructures de recherche et centres de données, producteurs de données, s'organisent pour répondre aux utilisateurs. Quelle est la stratégie de réponse aux utilisateurs ? Faut-il faire des recommandations à AERIS pour mieux répondre et anticiper la demande utilisateur ? Comment promouvoir les données de la recherche auprès du grand public et des décideurs et élargir les utilisateurs AERIS ?



La première question à se poser c'est qui est l'utilisateur d'AERIS ? Souvent les utilisateurs d'AERIS sont également des producteurs de données (même si certains jeux de données attirent des utilisateurs extérieurs comme les données IASI par exemple). Il faut également réfléchir aux utilisateurs hors domaine atmosphériques (domaine de santé par exemple) et **présenter les données de la façon la plus qualifiée et homogène possible pour qu'elles puissent être utilisées par des utilisateurs non expert du domaine**. Il y a également un enjeu lié à l'émergence de nouveaux acteurs qui vont distribuer et valoriser massivement des données. La plus-value des pôles de données est donc de présenter des données plus qualifiées que les fournisseurs commerciaux.

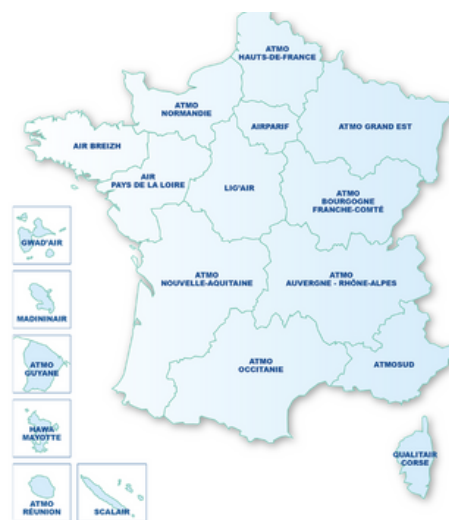
Il faudrait qu'AERIS soit complémentaire à Copernicus pour **présenter leurs données avec des exemples d'utilisation pour tout public** (mais cela demande un effort considérable et une documentation conséquente pour accompagner les utilisateurs sur l'utilisation des données).

Il y a une volonté dans Data Terra de développer des Jupyter Notebook pour **présenter les données par des exemples** (AERIS peut s'inspirer des NCAR data guide également).

Pour valoriser les données il y a un important travail à faire de communication, mais cela demande des moyens et une volonté d'acculturation des scientifiques. Il faudrait **élargir la communication aux écoles thématiques et aux étudiants. Il faudrait établir des indicateurs de comparaison sur des portails de données étranger pour pouvoir s'en inspirer pour la valorisation des données.**

Point important : AERIS est largement connue au sein de la communauté scientifique française du domaine atmosphère (grâce au soutien de nos tutelles). Mais il faudrait **élargir cette communauté utilisatrice** : il faudrait être **associé aux actions de communication des projets européens où AERIS est impliqué**. Il faudrait également faire un livret d'accueil pour présenter toutes les données d'AERIS qui serait relayé par les IR d'observation par exemple aux nouveaux arrivants. Il faudrait faire une veille scientifique des occasions de valorisation : pour aider à la mise en forme et publication des données et être visible (comme PANGEA) ?

Il y a un **lien à établir entre AERIS et les agences de qualité de l'air** : il y a une volonté d'établir des passerelles entre le monde opérationnel et le monde de la recherche grâce à la nouvelle politique de qualité de l'air et ils sont demandeurs de l'expertise scientifique sur les réseaux sol. Il y a effectivement des manques dans AERIS pour avoir ces données de qualité de l'air mais c'est possible maintenant d'obtenir ces données par des projets. Certaines personnes dans les AASQA sont sensibles à la recherche et le dialogue pourrait être facilité.



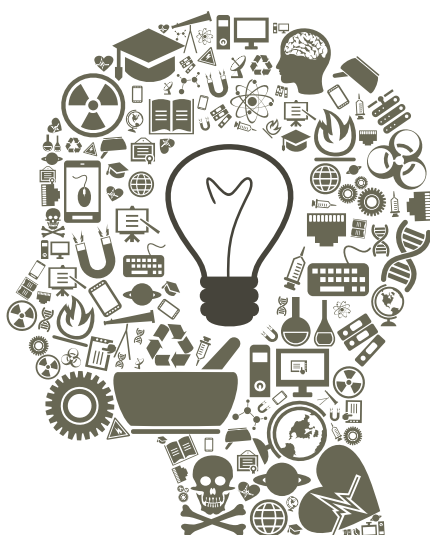
Carte des agences de qualité de l'air en France

AERIS essaie de valoriser les données par Data Terra et les autres pôles avec la **création de portails thématiques** (comme par exemple le portail urbain qui permettra de croiser les données d'hydrologie, données d'îlots de chaleur urbain...).

Il faudrait préciser le **cadre d'utilisation des données par les entreprises** : un travail est à faire sur les licences, mais le mandat du pôle n'est pas d'aller vers les entreprises.

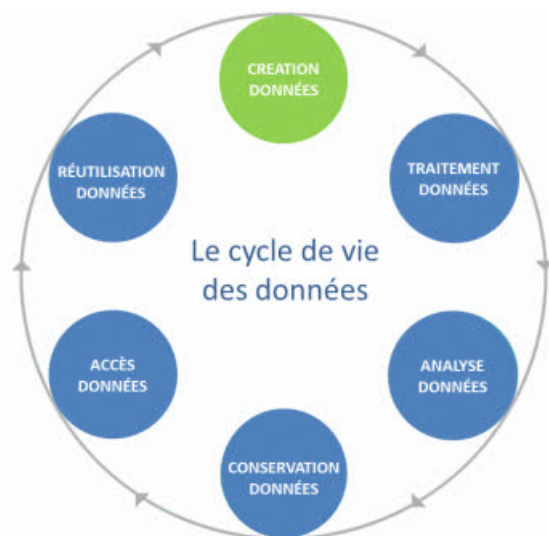
Il faudrait **utiliser le conseil scientifique pour faire le lien entre AERIS et les utilisateurs** : le mandat du conseil scientifique devrait évoluer et il devrait représenter toutes les communautés utilisatrices d'AERIS.

Un exemple à valoriser est le **soutien de la région Hauts-de-France depuis 2003 à AERIS** via son soutien à ICARE : une bonne dynamique est créée entre la région, ICARE et le LOA et il faudrait peut-être étendre cet exemple à d'autres régions. Il faut aller vers les élus, un discours est à mettre en place également sur les retombées (emploi, économique...) en plus de l'importance d'information des élus pour la prise de décision éclairée. Un enjeu connexe : Comment restaurer la confiance des citoyens dans la science ?



Atelier 5 : Ouverture et accès aux données : pourquoi, comment ? (Christelle Pierkot, Hélène Bressan, Karim Ramage)

Cet atelier est consacré au cycle de vie de la donnée et aux principes FAIR (Findable, Accessible, Interoperable, Reusable) : à chaque étape de ce cycle de vie, il est important d'avoir des retours et des besoins sur l'expérience des utilisateurs et producteurs de données d'AERIS et de voir les services d'AERIS/Data Terra associés (identifier les manques).



Cet atelier s'insère dans le contexte de la science ouverte : l'ouverture des données permet de faciliter la réutilisation des données par les chercheurs et évite les redondances. L'ouverture des données implique la documentation et la distribution des données dans des catalogues et entrepôts spécialisés pour permettre la réutilisation des données.

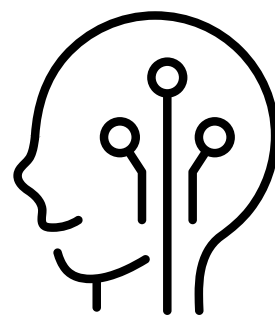
- **Création/Collecte** : Un exemple de retour est fait par les producteurs de données IASI : les données brutes L1 sont produites pour faire des L2 qui sont mises à disposition dans AERIS. Mais il y a un **manque de connaissance sur les vocabulaires à mettre en place pour la rédaction des fiches de métadonnées**. Il faut utiliser les standards ISO 19115 (utilisation pour les données géoréférencés) pour des métadonnées de découverte qui sont de plus haut niveau : qui a produit, licence, thématique, ... Il manque un **guide de rédaction des fiches de métadonnées** (AERIS accompagne au cas par cas les utilisateurs). Il faut mettre à disposition et expliciter pour les porteurs de projets des **éléments pour la rédaction des plans de gestion de données** (exemple sur la mise en œuvre de la pérennisation par AERIS ou Easy Data). Il y a un besoin d'**outils pour créer automatiquement des fiches de métadonnées à partir du contenu du fichier de données** (si ils sont dans des formats standards, comme NETCDF).

- **Traitement et analyse** : il faut préciser quel traitement, quel **code** a été utilisé pour produire la donnée. Il n'y a pas encore de bonnes pratiques dans AERIS sur ce sujet. Il faudrait s'inspirer de l'initiative Software Heritage soutenu par Recherche Data Gouv pour avoir un **catalogue des codes et logiciels** pour les référencer (niveau national). L'entrepôt EASY DATA est dans cette perspective également pour harmoniser et référencer les codes et logiciels.
- **Conservation** : quelles données doivent être pérennisées et dans quel entrepôt ? Il y a un **besoin de clarification entre Easy Data, Recherche Data Gouv et le catalogue AERIS**. Il faudrait que AERIS ou Data-Terra propose un petit texte à fournir aux dépositaires de données pour l'intégrer dans leur PGD. Il faut éviter de déposer des données dans Zenodo car il n'y a pas de curation de données dessus.
- **Accès** : Les producteurs de données expriment le souhait de pouvoir connaître les **statistiques d'accès et de téléchargement de leurs données** et de pouvoir prévenir les utilisateurs de leurs données de mise à jour ou autre événement en lien avec leurs données. Concernant l'accès aux données, la tendance serait **d'amener le traitement au plus près des données** et ne plus rapatrier les données chez soi, surtout pour les données volumineuses (comme les futures missions spatiales). Mais c'est compliqué pour les données étrangères (par exemple, les données GIIRS en Chine). Il faudrait qu'AERIS accompagne les chercheurs pour utiliser l'infrastructure technique des centres cloud thématique comme le EWC et sur l'utilisation de nouvelle API d'accès
- **Découverte et réutilisation** : Il faut bien choisir la licence que l'on souhaite pour ses données : les licences ne peuvent pas être changées après l'obtention d'un DOI. Concernant la découverte dans le catalogue AERIS, il faut **améliorer l'ergonomie de recherche du catalogue AERIS**. Par exemple l'utilisateur peut avoir du mal à trouver la donnée pertinente ou le producteur lui-même ne retrouve pas ses données.

Atelier 6 : Comment AERIS pourrait accompagner l'utilisation de l'IA ? (Sabine Schmidt, Vincent Martin, Nicolas Viltard)

Cet atelier vise à réfléchir sur comment mettre en œuvre des méthodes différentes sur l'IA et comment AERIS peut être force de proposition. Quels sont les usages de l'IA dans le domaine de l'atmosphère ?

Pour pouvoir utiliser des modèles IA il faut des données le plus **homogènes possibles (FAIR IA) avec le bon vocabulaire et le plus qualifié possible**. Il y a un enjeu de l'IA pour les **données de longues séries temporelles** : ces outils peuvent aider à dégager des tendances et faire des analyses multi-variables (ce qui peut être utile pour les données des réseaux de capteurs par exemple).



La réflexion sur l'IA ne touche pas que AERIS, elle doit être **mené au niveau de Data Terra**. AERIS peut en revanche avoir un rôle pour **qualifier et labelliser ses données et les rendre les plus compatibles aux méthodes IA** ainsi qu'à la préparation aux benchmarks pour développer les techniques IA (jeu de données complet, documenté pour permettre de faire de l'apprentissage). A terme, AERIS devrait donner accès à des données autre qu'atmosphère et échanger des données entre centres thématiques de Recherche Data Gouv. Des cas d'usage sont préparés dans Data Terra pour répondre à ces problématiques IA (un chatbot est en train d'être préparé dans le cadre de GAIA DATA avec des données du pôle ODATIS).

Il y a un important travail à faire dans la **documentation des données** pour pouvoir les utiliser dans des modèles IA. Par exemple, dans le cas de l'explosion du gazoduc Nord Stream, des données journalistiques (actualités) ont été croisées avec des données géophysique pour la détection d'anomalie. C'est également le cas avec l'imagerie où des données journalistiques sont comparées à des données d'imagerie satellitaire.

Enfin, il est important pour AERIS et Data Terra de **s'appuyer sur l'expérience technique de nos tutelles** avec notamment les travaux menés au CNES dans le campus de la donnée, ou au CNRS et à Météo-France (des contributions techniques du CNES sur ces sujets sont déjà données dans le cadre de GAIA DATA).