

**RAPPORT DU
GROUPE DE REFLEXION
« POLES THEMATIQUES EN OBSERVATION DE LA TERRE »**

AVRIL 2014

Ce rapport a été rédigé par le *groupe de réflexion « Pôles thématiques en observation de la Terre »*, mis en place par le Centre National d'Etudes Spatiales et l'Institut National des Sciences de l'Univers du Centre National de la Recherche Scientifique, et composé de :

- Gilles Bergametti (LISA et président du groupe TOSCA du CNES)
- Philippe Bertrand (INSU)
- Michel Diament (INSU)
- Jean-Pierre Gleyzes (CNES)
- Steven Hosford (CNES)
- Nicole Papineau (IPSL)
- Alain Podaire (CNES), animateur du groupe
- Marc Pontaud (Météo-France)
- Didier Roumigières (CNES)
- François Vial (INSU)
- Jean-Pierre Wigner (INRA)

Françoise Genova (INSU) a été associée aux travaux du groupe au moment de l'élaboration des propositions.

Mandat, périmètre et méthode de travail du groupe de réflexion

Mandat du groupe de réflexion

Le groupe de réflexion inter-organismes sur les pôles thématiques en observation de la Terre a été mandaté par le Centre National d'Etudes Spatiales (CNES) et l'Institut National des Sciences de l'Univers (INSU).

Le mandat du groupe de réflexion (voir annexe 1) précise que le concept de *pôles de compétences thématiques* – dits *pôles thématiques* – en observation de la Terre a été défini à la fin des années 1990, en s'appuyant sur les propositions contenues dans le *Rapport Waldteufel*.

Les *pôles thématiques* créés suite à ce rapport répondaient principalement au besoin de la communauté de recherche française de disposer de capacités de développement et de production de produits dérivés d'observation de la Terre depuis l'espace, et en particulier pour les missions définies et opérées par le CNES, en se plaçant dans une approche multi-capteurs.

Depuis la mise en place de *pôles thématiques nationaux* et d'autres structures comparables, de nombreuses évolutions ont eu lieu aux plans national, européen et international.

Parmi ces évolutions, on peut mentionner (i) des besoins scientifiques qui se sont affirmés, en particulier concernant l'accès à de nouveaux paramètres et aux séries temporelles nécessaires au suivi à long terme du comportement des compartiments terrestres, (ii) des évolutions techniques permettant de gérer de manière plus appropriée les volumes en forte croissance de données d'observation et produits dérivés, et (iii) le contexte et l'organisation aux niveaux national, européen et international, où notamment la demande de la société civile en information environnementale, et l'intégration de plus en plus nécessaire des capacités d'observation et de gestion des données au niveau européen sont essentiels et structurants pour les communautés de recherche nationales.

Ce nouveau contexte rend nécessaire une réflexion sur la stratégie, l'organisation et les actions nécessaires à moyen et long terme – donc dans une perspective postérieure à 2020 – pour, d'une part, satisfaire les communautés de recherche – et au delà les utilisateurs institutionnels et privés – au plan national et, d'autre part, participer à la coordination et l'intégration européennes et internationales à partir des compétences, des ressources et des infrastructures existant au niveau national.

Le groupe était mandaté pour conduire cette réflexion, au travers :

- d'une analyse de l'état des lieux national, européen et international, et des besoins à moyen-long terme des communautés de recherche,
- de propositions d'organisation pour les pôles thématiques et leur évolution, incluant les schémas de collaboration inter-organismes, une stratégie d'intégration européenne et internationale, et un plan d'action national.

Périmètre de travail

Le périmètre de travail du groupe de réflexion porte sur *la gestion et la diffusion des données d'observation et des produits à valeur ajoutée* – résultant de traitements qui sont appliqués aux données d'observation – *pour les géosciences*.

Par *données d'observation*, on entend :

- soit des données acquises par des satellites (*données spatiales*), de manière systématique ou après programmation (cas des données d'imagerie à haute ou très haute résolution en particulier),
- soit des données résultant d'acquisitions au sol, en mer, par avion ou ballon, que celles-ci soient acquises systématiquement (cas de réseaux ou de sites d'observation depuis le sol ou en mer) ou en utilisant des opportunités (campagnes, expériences diverses, ...) de durée plus ou moins longue.

Dans tous les cas, les données d'observation concernées se présentent sous un format numérique, et sont géo-localisées – c'est-à-dire repérables dans un référentiel géographique. Cela exclut donc du champ des travaux du groupe les données « matérielles » (carottes, etc.) ou non numériques, qui posent des problèmes de traitement et de conservation spécifiques non abordés ici.

Par ailleurs, les analyses et propositions présentées ici considèrent que les infrastructures – qu'elles soient spatiales, embarquées sur d'autres vecteurs, ou *in-situ* – nécessaires à l'acquisition des données d'observation sont en place.

L'existence de ces infrastructures d'observation constitue évidemment un préalable indispensable à la gestion des données qu'elles acquièrent. L'utilisation des données d'observation dépend de la couverture spatio-temporelle et de la précision des mesures collectées par ces infrastructures, mais également de la présence d'informations (métadonnées) suffisantes documentant les conditions dans lesquelles les observations ont été acquises et traitées.

Même si le développement et la mise en œuvre des infrastructures d'observation sont traités dans d'autres cadres d'analyse et de décision, et qu'en conséquence le groupe ne fera pas de propositions sur la nécessité de ces infrastructures ou sur leurs caractéristiques, le groupe se doit de souligner que la disponibilité en continu de systèmes d'observation appropriés, en termes de couverture spatiale et de fréquence et précision de mesure, est essentielle pour les activités de recherche ou de surveillance en support desquelles les systèmes de gestion de données interviennent.

*De plus, il est clair et admis que certaines activités liées à l'acquisition de données sont indissociables des activités de gestion des données associées, en particulier pour certains systèmes d'acquisition *in-situ*. Même si le groupe ne traite pas directement de l'acquisition des données, il est important de prendre en compte les mutualisations possibles (par exemple en termes de ressources humaines, d'infrastructures et de logiciels) entre les activités d'acquisition et de gestion des données.*

Enfin, les travaux de ce groupe ne traitent directement pas des développements, services et moyens liés à l'utilisation des données d'observation.

Les communautés utilisatrices des données d'observation ou des produits qui en sont dérivés sont au cœur de la gestion des données d'observation. Ce sont elles qui expriment des besoins (en données ou produits à valeur ajoutée), qui travaillent *sur* ou *avec* ces données, et qui donnent leur opinion sur l'utilité et la qualité des données et services de gestion des données.

Ces communautés sont donc – et doivent rester – pleinement impliquées dans l'expression de besoin et l'évaluation des systèmes de gestion des données. Et la structuration de la demande et des retours de ces communautés est sans doute déterminante pour la pérennisation des systèmes de gestion de données considérés.

Par contre, les moyens nécessaires à ces communautés pour exercer leurs activités, qu'ils concernent des ressources humaines ou des infrastructures (de calcul, d'archivage, etc.) ne sont pas directement pris en compte dans la réflexion présentée dans ce document.

Cette affirmation ne ferme pas les possibilités de mutualisation des moyens (ressources humaines, infrastructures) nécessaires à la gestion des données d'observation proprement dite avec ceux mis en

place pour la génération et la diffusion de **données simulées**, en particulier celles qui sont issues de modèles alimentés par les données d'observation.
Ces possibilités de mutualisation devront être intégrées dans l'analyse plus fine liée à la mise en œuvre des propositions du groupe,

En dépit de son intitulé et de son mandat, le groupe de réflexion a souhaité renommer son objet d'analyse, et employer le terme **pôle de données** de préférence celui de *pôle thématique*.
Ce changement de terminologie est principalement motivé par le fait que les missions et fonctions de gestion des données et de service traitées dans ce document sont mieux décrites par le terme pôles de données – qui peut se traduire en anglais par data center – que par celui de pôle thématique (lui même contraction du terme pôle de compétences thématiques).
De plus, les pôles de données gèrent des observations acquises sur les différents compartiments terrestres et des variables ou indicateurs qui en sont dérivés. Les pôles de données n'ont pas vocation à aborder des questions thématiques particulières, laissant ce soin à leurs utilisateurs.
La dénomination « pôles de données » permet donc d'éviter toute ambiguïté sur le périmètre de ceux-ci.

Méthode de travail

Le groupe de réflexion a commencé ses activités à la fin de l'année 2011. Il s'est réuni environ tous les deux mois lors de l'année 2012, puis à trois occasions en 2013.

Pour conduire sa réflexion et exécuter le mandat qui lui a été confié, le groupe a procédé en trois phases :

- une première phase d'analyse de l'état des lieux et des besoins des communautés de recherche : de janvier à avril 2012,
- une deuxième phase d'élaboration de propositions préliminaires : de juin à septembre 2012,
- une troisième phase de consultation et de finalisation des propositions : d'octobre 2012 à septembre 2013.

Lors de la première phase, le groupe a largement consulté les communautés scientifiques et acteurs nationaux et internationaux (en particulier en Allemagne, aux Etats-Unis et au Royaume-Uni). Compte tenu du large spectre d'activités et de thèmes couverts par son mandat, le groupe n'a pu être exhaustif et a procédé par échantillonnage, avec au final plus de quarante entretiens.

La deuxième phase du travail s'est déroulée de manière interne au groupe, en intégrant ponctuellement ou plus durablement des personnes expertes sur les différents aspects couverts par le groupe.

La troisième phase a consisté en une large consultation, basée sur les propositions préliminaires élaborées par le groupe, qui a permis, en intégrant les différentes remarques et suggestions, de finaliser les propositions et le rapport du groupe.

Ont notamment été consultés dans cette phase la plupart des organismes et institutions nationaux concernés par la gestion des données pour les géosciences, ainsi que les industriels impliqués directement ou indirectement dans ces activités.

Le 3 avril 2013, un atelier ouvert a rassemblé plus d'une centaine de participants, qui ont pu échanger avec le groupe sur la base d'un document rassemblant les propositions préliminaires de celui-ci.

La liste des personnes consultées figure en annexe 2.

Cette réflexion et le rapport qui la traduit ont été alimentés par de nombreux échanges avec des communautés et personnes motivées par la gestion des données d'observation de la Terre, que le groupe tient à remercier.

Le contexte et les acteurs

La gestion des données d'observation pour les géosciences se place nécessairement dans un contexte géographiquement ouvert, qui implique que le niveau national ne peut, la plupart du temps, être abordé indépendamment des niveaux européen et international, qui vont influencer et dans certains cas contraindre les activités et l'organisation nationale.

Toute approche ou stratégie nationale doit prendre en compte un contexte de plus en plus « globalisé ».

Le contexte et les acteurs nationaux

La gestion des données pour les géosciences au niveau national est liée, d'une part, aux systèmes d'observation opérés par les institutions nationales et, d'autre part, aux besoins des communautés nationales – de recherche et au-delà – en données d'observation, produits dérivés et services associés.

Les systèmes d'observation *nationaux* concernent soit des missions spatiales purement nationales ou développées en collaboration internationale, soit des systèmes d'observation *in-situ* ou sur vecteurs (avions, ballons, navires, ...) opérés sur le territoire national mais également hors de ce territoire, notamment lors de campagnes ou pour des contributions à des réseaux internationaux plus pérennes. L'existence de ces missions spatiales ou systèmes d'observation peut être liée à des besoins institutionnels hors recherche – les communautés de recherche se placent alors comme utilisatrices des données de ces systèmes – ou être définies en fonction des besoins et des spécifications des communautés de recherche – les communautés de recherche qui les ont définis sont naturellement les premiers utilisateurs de ces systèmes. Dans tous les cas, des systèmes de gestion des données sont associés aux infrastructures d'observation considérées.

Les données d'observation acquises par les opérateurs nationaux ne sont pas les seules à être utilisées par les communautés nationales des géosciences. Celles-ci s'approvisionnent aussi très largement en données d'observation collectées et délivrées par les opérateurs non-nationaux.

Cet approvisionnement multi-sources devient de plus en plus *structurel* à cause de la multiplication des sources de données, mais également du fait que les communautés de recherche s'intéressent de plus en plus aux dimensions continentales ou globales du système Terre et doivent générer ou utiliser des produits élaborés combinant plusieurs sources de données.

L'accès aux données internationales, mais aussi le développement de systèmes ou services permettant de les utiliser au mieux, sont donc nécessaires aux communautés nationales.

Les acteurs nationaux liés à la gestion des données sont :

- d'une part, les opérateurs institutionnels – ou plus rarement privés – des systèmes d'observation ; ces opérateurs peuvent avoir une mission d'observation ou de surveillance de la Terre – et souvent de service dérivé – pour le compte des institutions nationales, dont les institutions de recherche peuvent faire partie. Certains de ces acteurs exercent leur activité

outré-mer ou hors du territoire national – notamment pour des actions de développement avec des pays partenaires.

- d'autre part, les communautés utilisatrices, qu'elles conduisent des activités de recherche, de transfert entre recherche et services opérationnels, ou de surveillance du système Terre.

Le CNES a joué et continue à jouer un rôle essentiel dans le paysage national, à la fois par les missions qu'il développe et opère pour le compte de la communauté des géosciences – ce qui inclut une composante de traitement et gestion des données – mais également comme structure nationale travaillant en coopération avec les agences et opérateurs spatiaux au niveau européen (notamment ESA et EUMETSAT) et international. Ces activités en coopération incluent nécessairement un volet relatif au traitement et à la gestion des données acquises depuis l'espace.

Certains opérateurs privés, dont Airbus Defense and Space, sont également fournisseurs de données spatiales, essentiellement pour de l'imagerie à haute résolution spatiale.

Les autres organismes nationaux, dont le BRGM, le CNRS et l'INSU, l'IFREMER, l'IGN, l'INRA, l'IRSTEA, Météo-France, le SHOM, ... couvrent une large palette de fonctions et de compétences liées au thème traité ici, allant de la collecte systématique ou opportuniste d'observations en particulier *in-situ* ou par vecteurs autres que spatiaux, au traitement et à la gestion de ces observations, à l'expertise sur les données et les produits dérivés, mais également à l'utilisation de ces observations pour des activités de recherche, de surveillance et/ou de prévision.

Par ailleurs, l'évolution de l'organisation nationale de la recherche tant en termes de pilotage que de financement (avec, pour ce qui concerne les géosciences, l'existence d'Allenvi et d'Ancre, la programmation et les financements de l'ANR, mais aussi la nouvelle place des universités dans le dispositif national) devrait permettre une meilleure coopération entre les différentes institutions et organismes, notamment pour des activités à portée nationale, et une meilleure structuration de la base locale de ces activités.

Hormis les observations acquises par les organismes ayant un mandat institutionnel assorti d'une obligation de continuité d'acquisition, les dispositifs d'observation *in-situ* pour les géosciences sont structurés au niveau national autour de Services Nationaux d'Observation (SO), labellisés et pilotés par l'INSU, et dont certains ont vocation à assurer des observations systématiques, et de Systèmes d'Observation et d'Expérimentation au long terme pour la Recherche en Environnement (SOERE), qui mettent en réseau des observatoires de l'environnement déployés sur des sites différents, et dont AllEnvi assure l'évaluation, la structuration, la labellisation et le suivi. Par ailleurs, des observations peuvent être acquises en dehors d'un cadre ou d'un pilotage national, par exemple au niveau d'un laboratoire ou d'un Observatoire des Sciences de l'Univers (OSU).

Les dispositifs mentionnés ci-dessus n'ont aucune obligation contractuelle pour assurer la continuité à long terme des observations. Ils sont, par contre, généralement dotés de systèmes de gestion de données (bases ou centres de données).

En France, les communautés de recherche, qu'elles interviennent dans la collecte et la gestion des données d'observation ou dans l'utilisation de ces observations ou de produits qui en sont dérivés, sont structurées de manière « ascendante » sur plusieurs niveaux :

- le chercheur, ingénieur ou technicien pris individuellement, et l'équipe ou le laboratoire auxquels celui-ci est rattaché : cela concerne le *niveau local*,
- la fédération de recherche, de type Observatoire des Sciences de l'Univers (OSU), université ou institut, sur laquelle s'appuie l'équipe ou le laboratoire : cela concerne un *niveau de proximité, local ou régional*,

- des réseaux *nationaux*, par exemple entre plusieurs OSU, sur lesquels peuvent s'appuyer des infrastructures d'observation de type Infrastructure de Recherche – ou Très Grande Infrastructure de Recherche,
- les organismes ou institutions *nationaux*, quels que soient leurs statuts, qui contribuent au pilotage scientifique d'ensemble et à la fourniture de moyens.

A cette présentation simplifiée correspondent différents niveaux de mutualisation des moyens – du décentralisé localement au centralisé nationalement – avec des mécanismes d'attribution de ces moyens à la fois spécifiques et complémentaires.

Les propositions de ce rapport prendront nécessairement en compte les mandats et activités établies des opérateurs nationaux. Elles viseront notamment à ne pas dupliquer les activités et services déjà assurés par ces opérateurs, quand ceux-ci sont conformes aux besoins et exigences des communautés de géosciences.

Le contexte et les acteurs européens

Le niveau européen devient primordial pour les activités de gestion des données en général, et pour les géosciences en particulier.

Cette affirmation européenne passe en particulier par un rôle accru de l'*Union européenne*, tant sur les aspects réglementaires – qui peuvent se traduire par des contraintes sans contrepartie financière au niveau national – que programmatiques pour la recherche en géosciences et la surveillance de la planète Terre.

Concernant les aspects réglementaires pertinents de l'Union européenne, on peut notamment mentionner les directives INSPIRE et sur la *réutilisation des informations du secteur public* (Public Sector Information), qui définissent les mécanismes et méthodes pour documenter et accéder à l'information financée sur des budgets publics. L'objectif politique de cet ensemble réglementaire est d'ouvrir le plus possible l'accès aux données institutionnelles acquises dans chaque pays, afin que les institutions européennes et nationales puissent y accéder facilement. Cette facilité d'accès devrait développer l'utilisation de ces données et, si possible, de favoriser le développement d'applications marchandes les utilisant.

L'initiative GMES, devenue Programme Copernicus, constitue en termes d'objectifs et de financement, un élément structurant majeur du paysage européen de l'observation et de la surveillance de la Terre.

Copernicus repose sur des *services de base* fournissant de manière opérationnelle et pérenne des informations sur l'atmosphère (le *temps chimique*), les océans, les surfaces continentales, le climat, les risques et la sécurité. De ce fait, Copernicus permet de traiter – au-delà de domaines déjà structurés comme la météorologie – deux éléments : (i) l'intégration et la mutualisation, au niveau européen, de services publics d'information sur la planète Terre et, par conséquent, (ii) la continuité des observations, que celles-ci soit acquises depuis l'espace ou *in-situ*, nécessaires à la fourniture de ces services.

Ce Programme concerne au premier chef la communauté nationale des géosciences, à la fois parce qu'*une partie* des recherches qu'elle conduit trouve une finalité dans le développement et l'évolution de services publics européens (et nationaux), tant sur les méthodes que sur l'évaluation de la qualité, mais aussi parce que cette communauté est utilisatrice – et fait donc partie des *prescripteurs* – de ces services. De ce fait et à ce double titre, la stratégie nationale de gestion des données pour les géosciences doit s'intégrer au mieux dans le paysage européen dessiné par Copernicus.

Concernant la composante « *observation in-situ* » de Copernicus, il faut noter :

- d'une part, que l'Agence européenne de l'environnement (EEA) s'est vue confier, pour le compte de l'UE, le rôle de coordinateur de cette composante : ce rôle s'est essentiellement traduit par de la coordination politique et programmatique avec les Etats participants, sans assurer de fonction et de tâches de coordination technique. Par ailleurs, l'organisation de cette fonction après 2013, et notamment la pérennisation du rôle confié à l'EEA, restent à définir et décider.
- d'autre part, que cette composante repose essentiellement en termes d'infrastructures sur des contributions nationales, l'Union européenne se plaçant dans un rôle de coordinateur des systèmes nationaux nécessaires à Copernicus. La question de la contribution de l'Union européenne aux réseaux internationaux au travers de Copernicus fait débat et risque de ne pas être tranchée rapidement.

L'Union européenne pilote également des programmes cadres de recherche – dont Horizon 2020 pour la période 2014-2020 – qui incluent évidemment des activités sur les géosciences. Le programme européen d'infrastructures de recherche fait partie du programme cadre de recherche et de développement technologique et possède un volet sur les géosciences, qui concerne à la fois l'instrumentation – en particulier *in-situ* – et la gestion des données.

Toutefois, si le programme européen d'infrastructures de recherche est important pour mettre en place, sur la base de contributions nationales et en coordination européenne, des infrastructures d'observation et de gestion des données, il n'assure – du moins pour le moment – que dans certains cas seulement (dont la Terre solide) la pérennité des observations et donc la disponibilité *en continu* des données. De plus, les financements d'opportunité de tel ou tel programme de l'Union sont également inaptes à traiter cette pérennité.

Copernicus, avec son objectif de service s'inscrivant dans la durée, fournit un cadre programmatique, adéquat pour traiter de la pérennité des observations qui lui sont nécessaires, et en particulier les observations *in-situ* – principalement collectées à partir de systèmes nationaux.

Une nécessaire articulation, du ressort de l'Union européenne, sera néanmoins à trouver entre les programmes de recherche et Copernicus.

Pour ce qui concerne les observations depuis l'espace, deux acteurs principaux, tous deux régis par des accords intergouvernementaux, sont en place au niveau européen : il s'agit de l'Agence Spatiale Européenne (ESA) et d'EUMETSAT.

L'ESA est l'entité de niveau européen pour le développement et l'opération des missions scientifiques, en particulier en observation de la Terre et au travers du programme Earth Explorer. Ces missions sont essentielles pour définir, développer et valider dans l'espace des concepts et technologies d'observation. Par contre, Earth Explorer n'est pas le cadre programmatique approprié pour assurer la continuité des observations une fois les techniques d'observation validées.

C'est dans le cadre de Copernicus, auquel l'ESA a activement participé à la définition et à la mise en œuvre, que se joue la continuité d'une partie des observations spatiales européennes, en particulier pour l'imagerie optique et radar, les missions altimétriques et de chimie de l'atmosphère. L'ESA, auquel l'Union européenne a délégué la coordination du segment spatial de Copernicus, développe les missions spatiales de Copernicus – les Sentinelles – qui complètent les missions européennes et nationales existantes, notamment celles déjà opérées par EUMETSAT. Par contre, l'ESA qui n'a pas de mission – au moins pour le moment – d'*opérateur de continuité*, n'assure cette fonction *ad interim* que pour certaines missions Sentinelles qui n'ont pas d'opérateur public naturel.

En ce qui concerne la gestion des données spatiales issues des missions ESA, l'approche retenue consiste à moduler la décentralisation des fonctions de traitement et de distribution en fonction du niveau d'élaboration des données et produits : en gros, plus le niveau de traitement augmente, plus les fonctions sont décentralisées.

L'ESA est donc un interlocuteur naturel pour la communauté nationale des géosciences, pour laquelle les missions de recherche et opérationnelles développées en montage européen sont une source incontournable de données, intervenant en complément des missions gérées par le CNES.

EUMETSAT a un rôle d'opérateur européen en charge de l'acquisition, du traitement et de la diffusion des observations spatiales pour la météorologie et la climatologie. De ce fait, il opère ses propres missions – Météosat en orbite géostationnaire et EPS en orbite polaire – qui ont par nature vocation à assurer des observations systématiques et pérennes. Ces observations concernent les enveloppes fluides (atmosphère et océan), mais aussi les surfaces continentales.

EUMETSAT intervient également dans Copernicus, essentiellement en tant qu'opérateur de missions spatiales Sentinelles autonomes (Sentinelle 3 pour l'océan) ou cohabitant avec les missions EUMETSAT (Sentinelles 4 et 5 pour l'atmosphère).

Par ailleurs, EUMETSAT a mis en place un ensemble de huit Satellite Application Facilities (SAF) destiné, au travers de capacités distribuées dans différents pays, à générer et diffuser des produits à valeur ajoutée dérivés des observations EUMETSAT – et de certaines missions tierces.

Enfin, EUMETSAT a des accords de coopération avec plusieurs opérateurs de même nature au niveau international (dont la NOAA aux Etats-Unis) permettant d'une part de coordonner les systèmes d'observation et, d'autre part, d'échanger pour le compte de leurs utilisateurs respectifs les données de ces systèmes.

A ce titre, EUMETSAT est un fournisseur naturel de données pour les géosciences, en particulier pour ce qui concerne l'atmosphère, les océans et les surfaces continentales, et par suite pour les activités de surveillance climatique.

L'échelon national constitue un élément majeur de l'approche européenne pour la gestion des données d'observation et produits dérivés, en particulier parce que le niveau européen ne gère et ne finance pas la plupart des observations in-situ et la gestion des données associées, ni les traitements de niveau élevé ou très liés à un territoire national.

*C'est donc au travers d'une approche coopérative basée sur une logique de **co-production** entre acteurs nationaux et européens qu'il faut aborder l'organisation de la gestion des données pour les géosciences.*

Le niveau international

Les géosciences – et, au-delà, la surveillance de la planète – revêtent nécessairement une dimension globale, pour laquelle aucun pays ne peut prétendre à l'autonomie sur les systèmes d'observation, que ceux-ci soient spatiaux ou a fortiori in-situ.

Disposer d'un ensemble de systèmes qui observent de manière la plus exhaustive possible la planète et mettent à disposition de manière homogène leurs observations, relève nécessairement d'une coordination internationale basée sur un intérêt partagé, à la fois pour optimiser les capacités d'observation et pour échanger les données collectées.

La gestion des données d'observation est organisée au niveau international – hors Europe – au travers de coopérations bi- ou multilatérales, entre Etats ou fédérations, ou de coordination de niveau international s'exerçant dans un cadre Nations Unies ou apparenté.

L'intérêt principal de ces coopérations ou coordinations consiste à adopter des approches cohérentes en termes de systèmes d'observation à mettre en œuvre, en évitant si possible les duplications, et à favoriser l'échange et l'utilisation des données, notamment au travers de la mise en place de standards de format et d'accès aux données.

Parmi les cadres de coordination pertinents pour la gestion des données d'observation, on peut notamment citer le Conseil international pour la science (ICSU), l'Organisation Météorologique Mondiale, le Group for Earth Observation et son système de systèmes, le GEOSS, le Comité pour les Satellites d'Observation de la Terre (CEOS), qui concerne les agences spatiales et agit pour le compte du GEO.

Certaines coordinations s'exercent au travers de systèmes d'observation ciblés, tels que le Global Climate Observing System (GCOS), le Global Ocean Observing System (GOOS avec sa composante

européenne – EuroGOOS – et des structurations régionales), ou le Global Terrestrial Observing System (GTOS).

Les mandats et périmètres de ces différentes structures ont évidemment des intersections, ce qui complexifie l'organisation d'ensemble tout en favorisant des interactions – pas nécessairement efficaces.

De manière générale, les activités de coopération ou de coordination de niveau international se font sur la base d'accords ou d'approbations plus ou moins formels, dans une logique où chaque partie ou partenaire finance ses propres activités. Seules sont financées spécifiquement, en général par des contributions nationales ou communautaires de niveau peu élevé, les activités de coordination.

Les activités internationales s'exercent donc dans un cadre où les *meilleurs efforts*, les principes ou les bonnes pratiques, en général adoptés sans réel engagement, constituent la base de fonctionnement. Cette approche souvent informelle fonctionne convenablement, du fait de l'intérêt commun à bien se coordonner, et à échanger données et informations.

Certaines de ces structures accordent des labels – cas par exemple des Centres de données mondiaux (WDC) de l'OMM – qui permettent d'affirmer une position au niveau international. Il faut noter que très peu de centres de données français possèdent ce type de label.

Les activités de coopération internationale peuvent également se placer dans des cadres bi- ou multi-latéraux ; elle font en général l'objet d'accords avec engagements mutuels – de type *Memorandum of Understanding* – sur des projets ou programmes spécifiques, pour lesquels l'absence d'échange de fonds reste la règle. Par contre, la composante gestion des données d'observation est souvent intégrée dans un cadre plus large, lié au développement et à l'opération d'infrastructures d'observation.

Ce type de coopération est intéressant et doit être favorisé, car il permet de mettre en place de manière efficace soit des systèmes communs – les partenaires investissent sur un système unique – soit des systèmes complémentaires – on ne duplique pas ce qui est déjà traité par un partenaire.

L'échelon européen joue, par sa capacité de coordination mais également de financement, un rôle particulier dans l'approche internationale : la coordination européenne pourrait constituer un préalable efficace à un passage au niveau international.

Qu'est-ce qu'un pôle de données ?

Un pôle de données est un dispositif organisationnel et technique qui fournit à des utilisateurs un ensemble de services et d'informations. Le pôle se définit par la description des fournitures à destination de ses utilisateurs.

Les pôles de données s'adressent en priorité à la communauté scientifique française, mais n'excluent pas la possibilité de servir d'autres communautés d'utilisateurs (communautés scientifiques internationales, acteurs de politiques publiques, secteur privé, éducation et formation, etc.).

Les données d'observation sont considérées ici comme étant sous forme numérique. Les collections, carothèques, cartotheques, etc. ne sont pas concernées par cette réflexion, même si les pôles de données ont vocation à être associés aux dispositifs de conservation, d'archivage et de distribution des données sous une forme autre que numérique.

La mission d'un pôle de données : la fourniture de services

De manière générale, la mission principale d'un pôle est de produire et diffuser vers une large communauté d'utilisateurs des données, produits et services à valeur ajoutée ; le pôle s'efforce d'en faciliter l'utilisation et, par sa position centrale, en assure la visibilité et la promotion.

Les services d'un pôle incluent notamment – mais pas nécessairement :

- l'accès à des jeux de données qu'il détient et à des produits qu'il génère : ces jeux de données sont répertoriés dans un catalogue,
- l'aide à la découverte de données, à partir d'un guichet unique, et d'une sélection sur critères – l'accès à des catalogues de centres partenaires est possible si ces catalogues sont interopérables avec celui du pôle,
- la diffusion d'outils logiciels (e.g. visualisation ou intégration des données, changement de format, etc.) et de documents associés aux jeux de données ou au champ d'observations couvert par le pôle,
- un support d'expert sur les données et outils diffusés, afin de mieux les comprendre (performances, limitations, etc.),
- la fourniture de ressources pour, e.g. l'hébergement et l'exécution de codes de calcul, l'archivage de données, la publication de références, etc.,
- l'hébergement d'une communication éditoriale, basée sur des informations à portée générale ou plus spécifiques (travaux, projets, etc.),
- l'hébergement de plateformes collaboratives, de type Wiki, forum, blogs, production, partage et échange de documents...

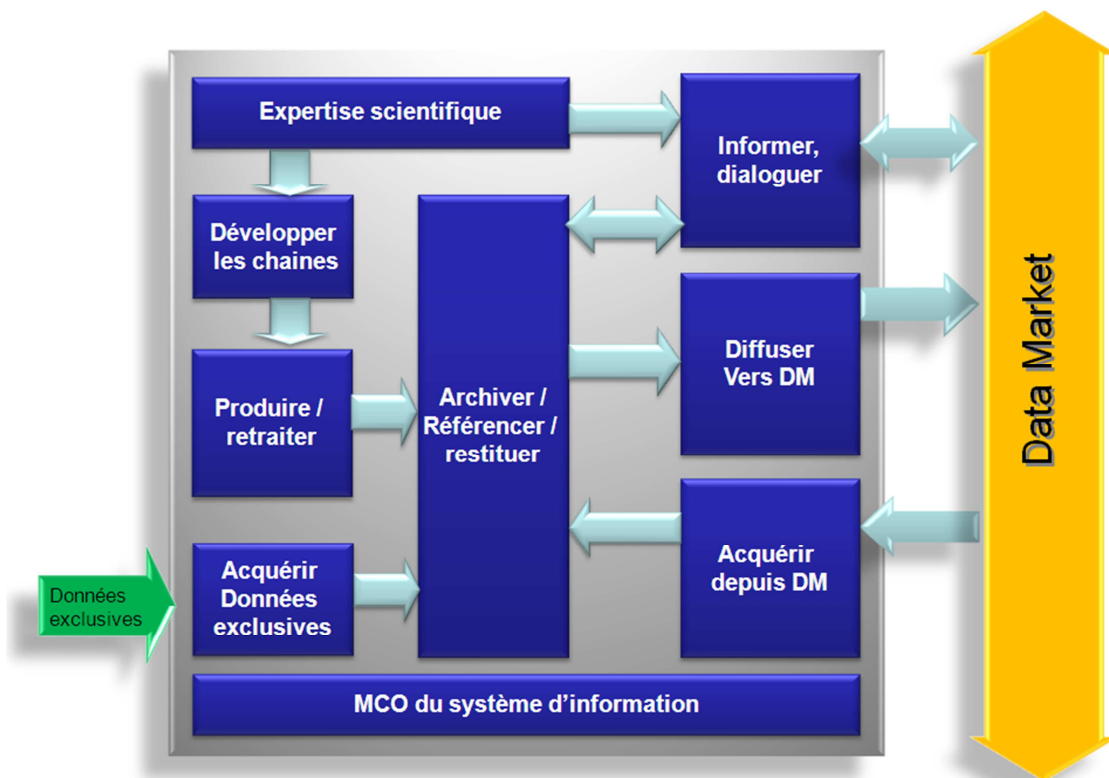
Les fonctions d'un pôle

Le pôle est structuré en un ensemble de fonctions internes, qui sont définies et organisées de manière à couvrir la fourniture des services à ses utilisateurs externes, et aussi à répondre aux besoins internes du pôle (gestion et entretien des moyens par exemple). *L'architecture fonctionnelle du pôle décrit cette structure, elle est commune à tous les pôles (des fonctions ou connexions peuvent être absentes dans certains cas) et ne dépend pas de l'architecture matérielle du pôle, qui est a priori spécifique de chaque pôle.*

La répartition des fonctions sur différents moyens ou lieux tient compte de contraintes spécifiques au pôle, comme par exemple la mise à profit d'un existant ou une répartition par *métiers*.

L'identification de fonctions communes à plusieurs pôles de données (ou communes avec d'autres entités partenaires) peut entraîner une démarche de mutualisation pour réunir des fonctions semblables sur des ressources et moyens partagés.

Le schéma fonctionnel d'un pôle peut être représenté comme suit :



Les principales interfaces d'un pôle sont :

- le « **Data Market** », qui représente tous les *media* auxquels le pôle est connecté pour échanger des informations avec d'autres centres de données ou avec ses propres utilisateurs. Les données échangées sur le Data Market ne sont pas exclusives, c'est-à-dire qu'elles ont vocation à une diffusion large, qui peut être restreinte dans certains cas.
- les **données exclusives** proviennent d'un dispositif d'observation extérieur, produisant des données uniquement à destination du pôle, et que le pôle est chargé de valoriser en réalisant des traitements à valeur ajoutée, leur archivage / pérennisation, leur diffusion / publication, et en fournissant un soutien d'expertise.

Les fonctions d'un pôle incluent :

- ***l'échange d'information*** : avec les utilisateurs (*informer / dialoguer*), le Data Market (*acquérir* auprès des fournisseurs extérieurs, *diffuser* vers le Data Market), les fournisseurs de données exclusives (*acquérir* des données auprès de ceux-ci),
- ***des fonctions internes*** : archivage/référencement/restitution, production/retraitement, développement de chaînes de production, organisation et fourniture d'expertise scientifique, production et maintenance d'outils logiciels, et maintien en condition opérationnelle (MCO) du système d'information.

Éléments d'architecture

Pour remplir les fonctions décrites précédemment, un pôle de données doit s'appuyer sur un ensemble de ressources (infrastructures, personnels, etc.). Ces ressources sont soit centralisées, soit distribuées géographiquement.

L'architecture physique d'un ***pôle de données*** inclut l'ensemble des ressources qui lui sont nécessaires, et leur géographie. Ce qui signifie qu'un pôle de données intègre généralement des fonctions centralisées et des fonctions distribuées géographiquement.

Ainsi, un pôle de données peut inclure dans son périmètre un certain nombre de ***centres de données***, c'est à dire des entités physiques concentrant en un même lieu des fonctions et des ressources contribuant à l'activité du pôle.

L'inclusion d'un centre de données dans le périmètre d'un pôle implique généralement des interfaces, règles et consignes de fonctionnement entre ce centre de données et les autres fonctions du pôle.

Éléments d'organisation

L'organisation d'un pôle de données est régie par un accord ou une convention liant l'ensemble des institutions – et le cas échéant des entreprises – contribuant, par fourniture directe de ressources humaines ou matérielles ou au travers de financements, au fonctionnement du pôle.

L'ensemble des moyens contribuant aux missions du pôle doit faire l'objet d'un contrat entre les différents partenaires du pôle.

Cela inclut notamment les fonctions d'expertise : sont considérées comme faisant partie du pôle celles qui s'appliquent aux produits et services fournis par le pôle, ou qui contribuent à préparer ses futurs produits et services.

Un pôle de données est doté d'une structure de décision, éventuellement assistée de comités d'experts, et d'une structure exécutive.

La structure de décision rassemble l'ensemble des partenaires signataires de l'accord ou de la convention constitutifs du pôle.

Les décisions prises par cette structure concernent la stratégie générale, les objectifs à court et moyen/long terme, les programmes de travail et les moyens attribués au pôle.

La structure de décision peut demander des conseils, évaluations ou propositions à des comités d'experts déjà existants et extérieurs au pôle, ou créer des comités permanents ou *ad hoc* pour l'assister dans l'exercice de son mandat.

La structure exécutive est composée de chercheurs, ingénieurs et techniciens couvrant l'ensemble des métiers nécessaires pour assurer la mission du pôle. Cette structure est mandatée par la structure de décision pour exécuter les programmes de travail que cette structure de décision a validés. Pour ce faire, elle gère les fonctions internes au pôle, les interfaces (techniques et scientifiques) nécessaires à

son fonctionnement, et les relations avec les utilisateurs du pôle. Elle rapporte à la structure de décision et participe, pour ce qui la concerne, aux propositions liées à la stratégie, aux objectifs et aux moyens du pôle.

Les besoins

Les recherches en géosciences sont pilotées par des **approches de plus en plus intégrées**, que celles-ci concernent le système Terre dans son ensemble, ou ses compartiments et les interfaces entre ces compartiments. La compréhension des processus et phénomènes reste une priorité majeure des géosciences, et la dimension temporelle – pour caractériser les évolutions, quantifier l'importance des phénomènes, etc. – est de plus en plus essentielle.

Les données d'observation et les produits qui en sont dérivés constituent évidemment un matériel capital pour ces recherches. Les besoins des communautés de recherche en géosciences traduisent ces tendances : ils sont multi-variables, multi-sources, et demandent souvent une profondeur temporelle importante.

L'objectif des pôles de données doit être de maximiser l'utilisation des données et produits d'observation. Ce qui implique que le *service rendu* par les pôles de données doit répondre aux besoins de *ses utilisateurs*, tout en apportant une *valeur ajoutée* par rapport aux produits délivrés par les structures extérieures aux pôles.

Les utilisateurs des pôles de données considérés ici sont les communautés de recherche nationales, avec une vocation à s'ouvrir aux communautés de recherche européennes ou internationales pour les composantes des pôles intégrées dans une configuration plus étendue.

Ce périmètre d'utilisation induit que la définition des produits (les catalogues) et services fournis par les pôles de données doit être pilotée prioritairement par les besoins des communautés de recherche.

Les communautés de recherche intéressées par les données et produits d'observation peuvent être schématiquement divisées en deux classes – très différentes, bien qu'un continuum et des échanges existent entre elles :

- les communautés qui s'intéressent à la technique d'observation, la physique du signal, l'étalonnage des observations et la validation des produits, etc, qui donc travaillent *sur les données*. Ces communautés ont été historiquement impliquées dans la mise en place des pôles ou centres de données, et sont généralement bien organisées pour exprimer leurs besoins en infrastructures d'observation ou produits dérivés. Les pôles de données doivent évidemment continuer à servir ces communautés.
- les communautés qui utilisent pour leurs travaux des observations ou produits dérivés, sans revenir sur la manière dont ils ont été générés ou validés. Ces communautés, qui travaillent *avec les données*, représentent un potentiel important pour développer l'utilisation scientifique des données. Leurs besoins, à la fois en termes de niveaux de produits (souvent plus élaborés que pour la communauté précédente) et de *niveau de service* (accès et utilisation plus simples, documentation, support expert) peuvent être sensiblement différents de ceux des communautés impliquées dans la physique de la mesure. A l'évidence, la disponibilité de données et produits d'observation pour ce type de communautés de recherche touchera également des communautés hors de la sphère recherche, qui peuvent utiliser ce type d'information pour des objectifs de suivi ou de gestion.

Les besoins des communautés des géosciences peuvent être déclinés en fonction des thématiques de recherche, mais également en termes d'organisation et de mutualisation des moyens en fonction des structures de recherche.

Les lignes directrices qui en découlent sont en particulier pilotées par la continuité des observations, qui permet de travailler sur de longues séries temporelles de données, et par la disponibilité d'observations plus ponctuelles ou d'opportunité liées aux études de processus.

La continuité des données d'observation : un besoin essentiel pour les géosciences

Depuis plusieurs décennies, les divers milieux terrestres subissent des évolutions significatives, en particulier en raison des conséquences des activités anthropiques. Cependant, compte tenu du rapport entre l'intensité de ces changements et l'amplitude de la variabilité naturelle du système Terre, dégager les tendances de ces évolutions nécessite d'effectuer des observations sur des périodes longues, au moins pluri-décennales.

Il existe donc, sur un certain nombre de variables clés, fortement indicatrices de l'état de la planète – niveau et couleur de la mer, extension de la glace de mer, salinité, contenu atmosphérique en gaz à effet de serre, champ magnétique, état de la végétation... – un fort besoin d'observations inscrites dans la longue durée, avec les gestions et archivages de données qui en découlent.

Pour les sciences de la planète Terre, le premier niveau de besoin exprimé par les thématiques de recherche est l'acquisition systématique de données d'observation de variables-clés des compartiments atmosphérique, océanique et continental.

Ce besoin, bien qu'en lisière du mandat du groupe, présente un fort intérêt pour la réflexion conduite dans la mesure où, d'une part ce besoin – en fait les inquiétudes liées à la prise en compte de ce besoin – a été exprimé à diverses reprises par les communautés auditionnées et, d'autre part, le volume des données d'observation qu'il induit va dimensionner les périmètres et les architecture des pôles et centres de données nécessaires.

Concernant les données *in-situ*, dans beaucoup de domaines, les systèmes d'observation (qu'ils soient labellisés ou non) ne font l'objet d'aucun engagement ou obligation contractuelle quant à la continuité sur le long terme de l'acquisition des observations : non seulement les agences et organismes peuvent suspendre leur financement, mais certains chercheurs en charge, n'étant parfois soumis à aucune obligation contractuelle, peuvent également arrêter à tout moment d'acquiescer des observations.

De même, au cours de ces quarante dernières années, le rôle au sein du dispositif des géosciences des données d'observation de la Terre depuis l'espace a profondément évolué, passant du statut de données *complémentaires* à celui de contributeur indispensable aux systèmes d'observation de l'environnement terrestre. C'est au niveau européen ou mondial que se joue la continuité des observations spatiales, et notamment en Europe au travers des programmes d'EUMETSAT pour la météorologie, l'océanographie ou le climat, ou de Copernicus pour un certain nombre de variables non couvertes ou complétant celles acquises via EUMETSAT.

Priorité absolue pour les géosciences, la continuité des observations in-situ et satellitaires n'est que partiellement acquise aujourd'hui, bien que diverses initiatives, notamment européennes et internationales, puissent y contribuer. Des capacités importantes d'archivage, d'inter-étalonnage, d'assurance qualité et de retraitement ou réanalyse des données doivent évidemment être associées à cette continuité.

Des données pour les études de processus

Outre l'étude des tendances d'évolution du système Terre, les progrès de notre connaissance et de notre compréhension des différents milieux et des liens/rétroactions existant entre eux constituent une priorité majeure pour les géosciences.

Les données d'observation sont également essentielles pour ce domaine de recherche, même si elles ont souvent une nature différente de celles nécessaires à la surveillance à long terme des milieux ; elles incluent notamment des données de campagnes (sol, aéroportées, océanographiques), des données

issues de vecteurs particuliers (ballons, planeurs sous-marins, voire éléphants de mer !), et des données d'expérimentation *in-situ* (mésocosmes, écotron...).

Ces jeux de données sont caractérisés par une fréquente limitation en *profondeur temporelle* (le plus souvent quelques semaines voire quelques mois) et par une grande diversité de paramètres documentés (l'intérêt des campagnes est de mesurer simultanément le maximum de paramètres d'intérêt sur un domaine géographique pour mieux comprendre le système et contraindre sa représentation numérique). On a donc des jeux de données de faible volume mais fortement hétérogènes.

On doit rajouter à ce type de données les données fondamentales de spectroscopie, de chimie, de réflectance, de biologie... obtenues en laboratoire.

Pour l'ensemble de ces données, le contexte international est souvent important car les grandes campagnes sont maintenant le plus souvent conduites dans un cadre bilatéral ou européen (voire plus large) et certaines bases de données (comme en spectroscopie ou en chimie) sont organisées au niveau international.

Des besoins en produits et services

Les entretiens et analyses qui ont été conduits font notamment ressortir les besoins génériques suivants :

- la nécessité de prendre en compte la demande des thématiques de recherche précisée plus haut, avec en particulier l'accès à des données et produits d'observation, multi-variables et multi-sources (acquises depuis l'espace, *in-situ*, ...), et la nécessité, pour qu'elles soient utilisables, que ces données soient bien qualifiées et documentées,
- la nécessité de couvrir une large gamme, du produit basique au produit élaboré, en fonction de la demande et de la qualification de l'utilisateur, et d'y accéder facilement,
- la mise à disposition de longues séries temporelles de données, dont la cohérence et la qualité doivent être quantifiées et documentées,
- l'importance de traiter les problèmes spécifiques liés à l'imagerie à haute résolution spatiale : licences, redistribution, synergie possible entre différentes utilisations,
- le « benchmarking » des produits des pôles et centres de données nationaux par rapport aux produits comparables existant ailleurs,
- l'harmonisation des standards (formats de données, métadonnées) pour la découverte, la visualisation et l'accès aux données, en lien avec les pratiques établies des communautés utilisatrices, et si nécessaire la prise en compte des contraintes réglementaires et principes mis en place au niveau européen et international,
- la mutualisation des catalogues, avec mise à jour dynamique ou plus simplement des liens entre les différents catalogues,
- la traçabilité des données, en termes de production et d'expertise,
- le besoin d'archivage de données, amplifié par la croissance des volumes et la diversité des variables et des sources de données à prendre en compte,
- l'intérêt de développer un lien entre observations et modèles, par exemple au travers d'analyses d'impact et de simulateurs d'observation, de type Observing System Evaluation ou Observing System Simulation Experiment (OSE/OSSE),

- le besoin de disposer des jeux de données « sur place » (près de l'utilisateur), et donc de les reproduire – sans toutefois dupliquer les fonctions qui demandent une sécurisation (archivage, etc).

Des besoins d'organisation et de mutualisation des moyens

Les besoins exprimés par les structures de recherche concernent à la fois l'utilisation et la production des données, souvent indissociables. Ces besoins portent plus sur l'organisation et la mutualisation des moyens associés à la gestion et à l'utilisation des données, que sur les données et informations elles-mêmes.

Si la gestion des données spatiales se traite souvent de *manière descendante* – avec un pilotage *par le haut* au niveau national et des ramifications locales – la gestion des données *in-situ* ou acquises/traitées localement est beaucoup *plus ascendante*, car fortement liée au système d'observation et à son opération, avant même d'envisager de faire remonter l'information.

Parmi les éléments les plus notables en termes d'organisation, on peut retenir :

- la reconnaissance, au niveau local et pour les observations *in-situ*, de la fonction et du rôle de **centre de données primaire**, qui assure la production de données et l'expertise sur les données produites, mais ne possède pas de mission obligatoire d'archivage et de diffusion de ces données. Ce besoin, formulé par les laboratoires ou les OSU, s'applique qu'un centre de données national existe ou non.
- la nécessité d'assurer une diffusion plus large (visibilité, accès, ...) aux données *in-situ* placées dans les laboratoires ou OSU, qui répondent bien aux besoins de recherche *locale*, et sont souvent gérées de manière hétérogène, non sécurisées ni documentées. L'adoption de standards pour les catalogues des bases de données locales pertinentes – assortie de certaines contraintes de normalisation et d'homogénéité – pourrait favoriser cette diffusion.
La complexité de certaines données (description et identifications taxonomiques, composition populationnelle, caractérisation écosystémique multi-paramètres, topographie, imagerie, etc...) impose que la structure productrice assure une fonction d'expertise sur ces données, et soit bien identifiée pour cela (traçabilité via un portail, un catalogue national de centre de données, ...).
- l'intérêt de gérer préférentiellement au niveau national l'archivage et la diffusion homogènes des mêmes types de données, et d'assurer l'interface avec des systèmes européens ou internationaux. Certaines structures de recherche (OSU notamment), ainsi que Coriolis pour certaines données océaniques, préfigurent déjà ce type d'organisation. On peut penser que, dans certains cas, la mise en place d'une base de données nationale mutualisée (hébergée dans un centre de données) rendra obsolètes les bases de données locales.
- la nécessité d'assurer et d'organiser une pérennisation (sécurisation) à long terme des bases de données. Dans la mesure où tous les centres de données (en particulier les centres de données primaires) n'assurent pas nécessairement cette fonction, il faut organiser des moyens permettant de gérer des bases sur des systèmes locaux et d'assurer leur archivage à long terme de manière plus centralisée au niveau national.
- la nécessité de réduire la segmentation des bases et centres de données, qui conduit à une dispersion des moyens et souvent à une duplication d'activité. Cette évolution permettra d'assurer un accès cohérent aux données/produits pour l'ensemble d'une thématique de recherche (correspondant souvent à une équipe, un laboratoire ou un programme national).

Elle doit également permettre la mise en commun de données issues de différentes communautés de recherche.

- Enfin, il faut très certainement construire des portails par compartiments du système Terre (atmosphère, océan, surfaces et interfaces continentales, terre solide,...) connectés avec des centres de données à différents niveaux, du national au local. Ce besoin provient de la communauté des géosciences, dans une vision multidisciplinaire et écosystémique où un chercheur peut avoir besoin de consulter les données et de bénéficier des compétences et de l'expertise existantes dans des champs très divers. Cela correspond aussi à un besoin des institutions de rendre globalement visibles, par grands secteurs thématiques, les données issues de leurs laboratoires et autres structures de recherche.

Il est important et nécessaire de ne pas dupliquer dans les pôles de données nationaux des activités et prestations qui sont déjà assurées par ailleurs, en particulier par des opérateurs institutionnels nationaux, européens ou internationaux. L'accès aux données et produits de ces opérateurs, nécessaire aux communautés de recherche, s'est fortement simplifié pour ces communautés, notamment après la mise en œuvre de directives européennes. Le référencement des portails de données de ces opérateurs devrait permettre de mieux informer sur les données et produits existants et sur les moyens pour y accéder.

L'existant

Bien qu'ayant largement consulté les entités et personnes impliquées dans la gestion des données d'observation au niveau national, le groupe de réflexion n'avait pas pour mandat ou objectif de réaliser une évaluation détaillée de chaque structure assurant une mission de ce type. Des organes et mécanismes d'évaluation sont en place pour ce faire. L'analyse et les constats reportés dans ce document sont donc généraux – et doivent être considérés comme tels – et ne traitent pas des spécificités de telle ou telle structure de gestion des données.

En France, la gestion des données d'observation pour la recherche en géosciences présente de grandes différences d'approches et de structuration, dépendant des communautés et disciplines concernées :

- certaines communautés/disciplines ont déjà mis en place leurs centres de données, que ce soit au niveau de structures de recherche et/ou d'observation (type OSU) ou au niveau national, et peuvent fournir des retours d'expérience sur l'existant. Ceci inclut des disciplines/communautés liées à des systèmes opérationnels et nécessairement structurées par ce lien,
- d'autres abordent la mise en place de leurs systèmes de gestion des données,
- d'autres en sont à définir leurs systèmes de gestion des données, et conduisent en premier lieu une analyse des besoins.

Dans tous les cas, il existe déjà des composantes d'un système ou un système complet, qui peut éventuellement évoluer. Comme l'existant repose évidemment sur des moyens (infrastructures et ressources humaines), l'analyse et la prise en compte « inclusive » de ces moyens est essentielle au démarrage et/ou à l'évolution d'un service.

Les approches de gestion de données retenues pour les observations *in-situ* – qu'elles se placent ou non dans le cadre de Services d'Observation – sont sensiblement différentes de celles des systèmes de gestion des données spatiales.

Compte tenu en particulier des volumes plus faibles générés par les systèmes d'observation *in-situ*, et souvent aussi de la diversité des observations, la gestion associée des données est plutôt décentralisée et organisée autour de l'entité qui opère le système d'observation. Ce qui n'exclut pas, notamment dans le cas fréquent de contribution à des réseaux internationaux, de garantir une fourniture de données conforme à une organisation et des standards (format, qualité ...) de portée plus large.

Pour les données *des missions spatiales*, la mise en place des centres de données relève plus d'une logique de projet, avec des phases de développement, puis d'exploitation et de maintenance des systèmes. Le centre de données est l'une des composantes du système mis en place pour chaque mission spatiale, avec toutefois – et de manière de plus en plus fréquente – la prise en compte d'éléments multi-missions/instruments.

D'une manière générale, on constate qu'un grand nombre de pôles ou centres de données existants fonctionnent de manière satisfaisante : ils fournissent un service apprécié par les communautés qui les utilisent et sont animés par des équipes motivées, malgré des moyens souvent limités.

Les questions principales qui se posent concernent l'évolution des pôles ou centres de données existants dans un contexte différent de celui dans lequel ils ont été pensés et construits, ainsi que les nouveaux pôles ou centres de données en cours de définition ou de mise en place.

Ces évolutions de l'existant, tout comme la mise en place de nouveaux pôles/centres de données, doivent être guidées par les constats généraux suivants :

- *les utilisateurs et leurs besoins* : certains pôles/centres de données existants ont été définis par/pour des communautés utilisatrices, selon les besoins exprimés au moment de cette définition. Il est nécessaire de réévaluer ces besoins en fonction (i) des utilisateurs qui sont visés (élargissement éventuel de la base d'utilisation) et (ii) des besoins actuels et futurs de ces utilisateurs (évolution des thématiques de recherche, utilisation de données d'observation d'origines variées, y compris en dehors des domaines de compétence de ces utilisateurs).
- *le développement des pôles/centres de données, et en particulier des produits et informations qu'ils génèrent et délivrent* : les systèmes destinés à servir les communautés de recherche en géosciences doivent faire preuve d'une grande souplesse, en particulier lors du développement et de la mise au point de chaînes de traitement – il ne faut pas consolider prématurément des configurations destinées à évoluer...
- *les opérations des pôles/centres de données, incluant l'évaluation de la qualité* : les objectifs et l'organisation des pôles et centres de données doivent prendre en compte l'évolution des communautés utilisatrices, en particulier les utilisateurs peu ou pas impliqués dans la conception et/ou la validation des produits. Par ailleurs, la demande croissante en séries temporelles requiert des capacités de retraitement et des activités de mise en cohérence et de qualification scientifique des séries d'observations.
- *l'accès aux données* : il doit être amélioré et tendre vers toujours plus de simplicité pour se procurer les données ou s'enregistrer en tant qu'utilisateur (moins de procédures ou formulaires), une meilleure description des produits disponibles (catalogues), des formats standardisés avec un support à l'utilisation (expertise au travers de documents ou de support direct).
- *l'archivage des données d'observation et produits dérivés* : les questions d'archivage patrimonial des observations se posent de manière urgente du fait de l'augmentation de leurs volumes et de leur diversité.
- *l'organisation des pôles/centres de données et les mécanismes de décision* : l'organisation doit être rendue plus efficace et lisible, afin de mieux caractériser les activités et les moyens qui y sont affectés, simplifier les processus de décision, faciliter les interactions entre le niveau de décision et l'exécution des tâches et améliorer la représentation dans les instances nationales et internationales.
- *les moyens affectés aux pôles et centres de données* : le manque de moyens, ou parfois la dispersion ou le manque d'identification claire de ceux-ci, fragilisent ou rendent difficile la réalisation de certaines fonctions techniques et scientifiques des pôles / centres de données.
- *l'intégration européenne (et internationale)* : elle doit être traitée au niveau scientifique et technique, mais également au niveau institutionnel, avec des engagements à long terme, en mettant en place des règles de réciprocité avec les partenaires européens ou internationaux (pas de duplication injustifiée des activités et moyens, utilisation des systèmes partenaires, ...).

Les perspectives à long terme et la stratégie nationale associée

La stratégie nationale de gestion des données pour les géosciences devrait avoir pour principal objectif de maximiser l'utilisation des données d'observation acquises sur financements publics, que ceux-ci soient nationaux ou européens, et de contribuer à la fois au progrès de la connaissance de la planète Terre et à la surveillance de celle-ci, au travers de travaux de recherche et d'utilisations répondant à des demandes sociétales.

Les perspectives à long terme pour les pôles de données nationaux doivent résolument se placer dans un contexte européen.

Les objectifs et approches pour les recherches en géosciences sont de plus en plus intégrés ; les moyens d'observation mis en œuvre au niveau européen reposent sur des contributions nationales, mais également sur des infrastructures directement développées et opérées au niveau européen (missions spatiales notamment).

Il est donc envisageable, et plusieurs signaux à la fois programmatiques et politiques/réglementaires tendent à le confirmer¹, que l'Europe mette en place – à terme – des moyens d'accès coordonnés à toutes les données d'observation acquises sur budgets publics.

Ce qui signifie que les données et produits d'observation pour la recherche, en particulier nationales, devront être répertoriées et accessibles au travers de ce système européen.

Compte tenu de ces perspectives à long terme, il convient de définir et mettre en œuvre une stratégie nationale autour de la gestion des observations pour les géosciences permettant, d'une part, de servir au mieux les communautés françaises concernées et, d'autre part, d'être acteurs à un niveau correspondant aux moyens et compétences mis en œuvre et à la stratégie nationale dans l'approche européenne d'intégration de ces activités.

L'intégration européenne repose sur des contributions naturelles de niveau national, qui concernent notamment les données acquises sur le territoire national.

Elle suppose également que des positions françaises seront prises sur des éléments bien identifiés de niveau européen, en particulier sur des fonctions mutualisées à cet échelon. Ces positions françaises doivent se situer dans une logique de réciprocité avec les partenaires européens, basée sur la mise en valeur des capacités et des domaines de compétence nationaux.

L'approche européenne va également demander de se configurer au niveau national pour être dans la meilleure position possible pour l'intégration à l'échelon européen. Et de définir les *trajectoires* à moyen et long terme pour aboutir à cet objectif.

De ce fait, l'organisation nationale des pôles de données et leur architecture, qui sera nécessairement distribuée, doivent être compatibles avec les fonctions et standards européens.

Cela suppose d'organiser les systèmes de gestion des données nationaux d'une manière cohérente et lisible au niveau international, en coordonnant techniquement leurs fonctions principales, et en ayant la capacité de négocier les positions nationales au niveau européen.

¹ Voir en particulier la directive INSPIRE, le Shared Environment Information System (SEIS), la mise en place de Copernicus, la consultation autour de Marine Knowledge 2020, les priorités affichées pour Horizon 2020...

Cela suppose également de *faire porter l'effort, au niveau national, sur certaines fonctions des systèmes de gestion des données* ; par exemple le développement de chaînes de traitement, la qualification des produits, la diffusion de données et produits, et l'information sur ceux-ci.

Un effort tout particulier doit être consacré, au niveau national, aux *fonctions et éléments qui vont devenir stratégiques dans les futures configurations européennes et internationales*. Ainsi, les *fonctions d'accès aux données* – notamment au travers de portails – ont été parfois délaissées alors qu'elles sont essentielles pour le contact avec les utilisateurs, la visibilité sur les données et sur les activités nationales, mais également pour l'intégration au niveau international.

La recherche de *labels internationaux* qui permettent, avec des engagements techniques souvent limités, d'être reconnu au plan international, est également à considérer.

Il conviendra de définir la place des opérateurs « recherche » en charge de la gestion des données par rapport aux opérateurs nationaux et européens mandatés pour générer et diffuser des données et informations pour des objectifs institutionnels plus larges que la recherche – mais incluant celle-ci.

Les données et produits pertinents des opérateurs institutionnels doivent être utilisables pour des besoins de recherche, et le sont souvent. A contrario, les opérateurs « recherche » ne peuvent assurer que certains *niveaux de service*, liés aux ressources et métiers dont ils disposent.

La préservation des compétences des opérateurs « recherche » nationaux, notamment autour de l'expertise scientifique, est un élément essentiel de la stratégie nationale.

Les systèmes nationaux de gestion de données pour les géosciences sont du ressort de la sphère publique. Les modèles économiques associés sont donc essentiellement basés sur du financement public national ou européen. Et ce, quels que soient les utilisations et utilisateurs des données et produits qu'ils distribuent.

Les politiques de données, en particulier européennes, autour des observations pour les géosciences acquises sur financement public s'orientent très clairement vers *leur mise à disposition libre et gratuite*. Certaines restrictions, liées à la qualification des données, à leur caractère confidentiel ou à des licences d'utilisation préexistantes, peuvent être prises en compte, sans toutefois remettre en cause le principe de libre accès.

De plus, les *nouvelles capacités technologiques* vont permettre de gérer de manière cohérente des configurations complexes, et vont accompagner la simplification et la banalisation de l'accès aux données d'observation, rendant caduques certaines segmentations actuelles entre les types et origines des données. Ces nouvelles technologies doivent être dès maintenant considérées dans les architectures des systèmes nationaux.

La *valorisation des données d'observation* passe par leur utilisation dans le domaine de la formation secondaire ou universitaire, mais également par la génération et la diffusion d'informations destinées au « grand public » – et facilement compréhensibles par celui-ci.

Ces approches de valorisation, importantes pour la visibilité de la société civile sur les activités de recherche liées aux observations, doivent être considérées comme prioritaires.

Par ailleurs, les efforts et moyens consacrés par les personnels des organismes pour le partage et la dissémination des données doivent trouver une reconnaissance au travers de nouvelles formules de publication et de référencement des jeux de données qu'ils ont constitués.

La perspective à long terme d'intégration européenne des systèmes nationaux de gestion des données pour les géosciences doit piloter la stratégie nationale sur l'évolution de ces systèmes. Elle implique d'abord une intégration progressive des systèmes existants au niveau national, prenant pour base les fonctions qui sont en place et sécurisées, et consolidant certaines fonctions insuffisamment développées ou assurées, et la mise en place ou la redistribution des moyens (infrastructures et personnels) correspondants.

Les propositions du groupe de réflexion

Propositions sur l'organisation nationale

Proposition n° 1 – La définition d'une stratégie nationale sur l'intégration européenne des pôles de données est un préalable

La définition d'une stratégie nationale sur l'intégration européenne des pôles de données est un élément déterminant de leur structuration.

Elle doit notamment permettre d'identifier et délimiter :

- *ce qui est fait au niveau national pour des utilisations limitées au national,*
- *ce qui est fait au niveau national pour des utilisations nationales et européennes,*
- *ce qui est fait au niveau européen et est utilisé au niveau national, et ne doit donc pas être dupliqué au niveau national.*

Cette stratégie doit être établie pour chacun des compartiments du système Terre (voir la proposition n° 2) à partir d'une approche d'ensemble et de principes s'appliquant à tous les pôles de données nationaux.

Elle doit permettre de définir les positions (fonctions) nationales relevant du niveau européen, en prenant en compte les contraintes (notamment de pérennisation de moyens mais aussi de conformité aux règlements et directives européennes) et les engagements institutionnels associés.

Elle ne doit évidemment pas remettre en cause les positions acquises au niveau européen, mais plutôt s'appuyer sur cette base.

Les liens entre les pôles de données et les programmes de l'Union européenne – dont Copernicus et les infrastructures de recherche – ainsi qu'avec les programmes et activités des agences spatiales comme l'ESA ou EUMETSAT doivent être traités dans le cadre de cette stratégie nationale.

La stratégie européenne proposée peut s'appuyer sur des coopérations renforcées avec certains pays européens, permettant notamment de répartir au mieux fonctions et moyens entre les institutions françaises et celles de ces pays, et de constituer un socle pour l'édifice européen.

Compte tenu du contexte européen autour des géosciences (notamment le démarrage d'Horizon 2020 et de la phase opérationnelle de Copernicus – y compris avec la mise à disposition de nouvelles données spatiales, telles que celles des Sentinelles – dès 2014), cette stratégie nationale doit être définie le plus tôt possible – avant mi-2014.

Cette stratégie européenne devra naturellement prendre en compte les articulations existantes ou à établir avec le niveau international, au delà de l'Europe.

Proposition n° 2 – Vers quatre pôles de données nationaux

Il est proposé de mettre en place, à terme, quatre pôles de données nationaux, soit un par compartiment du système Terre : atmosphère, océan, surfaces continentales et Terre solide.

Chacun de ces pôles devra prendre en compte l'ensemble des données et produits d'observation correspondant à son compartiment, que ces observations soient acquises depuis l'espace, à partir d'avions, ballons ou autres vecteurs aériens, au sol ou en mer...

La mise en place de ces quatre pôles de données répond à plusieurs besoins ou objectifs, à savoir :

- une meilleure réponse aux *besoins des utilisateurs nationaux et, le cas échéant, européens et internationaux*, pour accéder de manière aisée et cohérente à des données multi-sources et multi-variables, bien qualifiées et documentées,
- une *organisation technique et scientifique* – autour des fonctions et des moyens – *simplifiée*, permettant de mieux définir et utiliser les ressources, de constituer des masses critiques, de formaliser un *fonctionnement en réseau dans une logique de co-production*,
- *plus d'efficacité et de lisibilité dans les processus de décision* (en particulier sur les moyens), d'exécution et d'évaluation des activités, ainsi que de visibilité et de représentation aux niveaux national et international,
- une *organisation plus conforme à celle des structures similaires existant au niveau international*, et notamment au niveau européen, afin de favoriser l'intégration européenne des activités,
- une anticipation sur l'utilisation de nouvelles technologies (Big Data, data centres, cloud, etc.) qui vont améliorer certaines fonctions et simplifier les interfaces avec les utilisateurs (terminaux virtuels, découplage entre travail sur les données et stockage physique de ces données, ...).

La segmentation en quatre pôles nationaux de données correspond à une structuration des communautés de recherche sur les géosciences, que ce soit en termes d'utilisation, de gestion des infrastructures et d'expertise.

La constitution de chacun de ces quatre pôles de données se place dans une perspective à moyen terme, avec des échéances et étapes de structuration différentes pour chacun de ces pôles.

Elle ne remet pas en cause les structures de gestion de données existantes, mais vise au contraire à mieux les intégrer au niveau national, en évitant certaines duplications et/ou en renforçant les masses critiques de moyens sur certaines fonctions. Elle devra donc se construire à partir de ce qui existe et de ses contraintes.

Compte tenu des moyens mis en œuvre, les pôles nationaux de données pour les géosciences devraient être reconnus comme une (ou des) infrastructure(s) de recherche.

Proposition n° 3 – Une forte structuration inter-pôles

La proposition de constitution de quatre pôles de données nationaux – un par compartiment du système Terre – ne vise pas à segmenter de manière irréversible le paysage national de la gestion des données d'observation. Elle doit au contraire être vue comme une étape vers une intégration plus forte entre les pôles, pouvant éventuellement préfigurer la constitution d'un pôle national unique de données.

La structuration inter-pôles est en particulier nécessaire pour :

- traiter des besoins scientifiques se plaçant aux interfaces entre les compartiments : les activités de recherche portant sur les interfaces du système Terre prennent une importance de plus en

plus grande, et il est nécessaire que les données d'observation utilisées pour traiter cette thématique soient mises en cohérence entre les différents pôles de données ;

- prendre en compte les définitions de périmètres des pôles pour lesquels certains milieux – comme l'hydrosphère – pourraient être traités au travers de plusieurs pôles ;
- réaliser des développements techniques – par exemple sur certaines fonctions de diffusion ou de traitement de l'information – pour le compte de l'ensemble des pôles ou partageables entre ceux-ci ou, au minimum, avoir une plate-forme d'échanges et de partage d'informations techniques ;
- assurer un pilotage programmatique et stratégique cohérent pour l'ensemble des pôles nationaux, qui prenne en compte les dimensions nationale, européenne et internationale.

Il est proposé que, dans un premier temps, cette structuration inter-pôles se réalise au travers d'un pilotage programmatique et stratégique cohérent pour l'ensemble des pôles de données nationaux et de la mise en place d'organes permettant de traiter transversalement les aspects techniques pertinents (voir proposition n° 4).

Proposition n° 4 – Une gouvernance et un exécutif renforcés et plus lisibles

Chaque pôle de données doit pouvoir travailler de manière autonome, structurellement et administrativement. Ce qui implique qu'il définisse ses objectifs et activités, que ses processus de décision soient spécifiques et qu'il dispose de moyens dédiés à l'exécution de ses activités.

Il est proposé que chaque pôle dispose **pour sa gouvernance** (ses processus de décision) :

- **d'un comité directeur** composé des institutions contribuant aux moyens (ressources humaines, infrastructures, etc.) du pôle, avec pouvoir de décision sur le plan d'activité du pôle et son approche de mise en œuvre, ses moyens, sa stratégie nationale, européenne et internationale,
- **d'un conseil scientifique**, nommé par le comité directeur, composés d'experts sur les données et produits du pôle, et de représentants des programmes de recherche ou institutions nationales pertinentes, qui évalue les activités du pôle et les demandes qui lui sont faites, et contribue à la préparation de la stratégie de celui-ci.

Il est proposé que **l'exécutif d'un pôle se compose** :

- **d'un directeur**, nommé par le comité directeur du pôle, qui est en charge de la mise en œuvre des activités du pôle, de la préparation de sa stratégie et de sa représentation auprès des instances nationales, européennes et internationales. Il coordonne l'ensemble des fonctions du pôle et est responsable de son organisation et du bon fonctionnement de ses moyens, et de la gestion de ses personnels. Il préside le comité scientifique du pôle. Il rapporte sur l'ensemble des activités au comité directeur.
- **d'équipes techniques et scientifiques**, assurant l'ensemble des fonctions du pôle de données, si nécessaires distribuées géographiquement, et coordonnées par le directeur du pôle.

Il est par ailleurs proposé que **des instances inter-pôles soient mises en place**, afin de définir et de mettre en œuvre une stratégie nationale cohérente et concertée, incluant des activités techniques communes.

Ces instances pourraient comprendre :

- **un comité de pilotage inter-pôles** composé de représentants des organismes contribuant aux moyens des pôles de données et des directeurs des pôles, en charge de définir et suivre la mise en œuvre de la stratégie nationale,

- **un comité technique inter-pôles**, composé des directeurs des pôles ou de leurs représentants, ainsi que de représentants d'organismes, en charge de définir les développements, règles ou consignes techniques permettant d'assurer une cohérence entre les pôles et la réponse aux besoins *transverses* des utilisateurs.

Cette proposition concernant la gouvernance et l'exécutif des pôles de données doit permettre une meilleure visibilité sur les activités et les moyens de ces pôles, ainsi qu'une représentation améliorée de ceux-ci dans les instances nationales et internationales. Elle répond au constat général de dispersion et de manque de cohérence ressortant de l'organisation actuelle.

Sa mise en œuvre impose de résoudre des questions spécifiques, notamment sur les rattachements et évaluations des personnels.

Proposition n° 5 – L'expertise scientifique au cœur des pôles de données

L'expertise scientifique constitue un élément essentiel des pôles de données. Elle s'applique à des fonctions telles que la définition des méthodes de traitement ou le développement et l'optimisation des chaînes de traitement, la qualification des données et produits des pôles – par exemple leur étalonnage et validation – qui doit se traduire par des documents ou des publications diffusables largement, le support aux utilisateurs, que ce soit sous forme documentaire ou en contact direct avec ceux-ci, mais également la formation et la communication.

Cette expertise concerne les produits diffusés par un pôle de données à un instant donné – qui doivent être répertoriés dans un catalogue – mais également l'amont pour préparer les futurs produits du pôle, soit par évolution des produits existants, soit par l'intégration de nouveaux produits.

L'expertise telle qu'elle est définie ci-dessus se distingue des activités – qui requièrent indéniablement une expertise – liées à l'utilisation des données produites par les pôles, qui ne sont pas directement contrôlées par ces pôles et ne participent donc pas directement aux activités de ceux-ci.

Il est souhaitable que l'expertise scientifique participant directement au fonctionnement et aux activités d'un pôle soit contractualisée par le pôle en question.

Cette contractualisation pourrait prendre deux formes :

- pour ce qui concerne les produits répertoriés à un instant donné dans le catalogue d'un pôle, l'expertise concerne principalement la qualification en continu des données et produits, et le support aux utilisateurs : il s'agit donc d'activités à *caractère systématique* que l'on peut aisément traiter au travers de contrats classiques spécifiant les prestations demandées ;
- pour ce qui concerne l'évolution des produits du pôle ou la création et la qualification de nouveaux produits, les activités associées relèvent plus directement de la recherche, soit à partir de prescriptions venant des utilisateurs et transmises par le pôle de données, soit sans prescription spécifique.

Il est souhaitable qu'une enveloppe financière spécifique, à chaque pôle ou pour l'ensemble des pôles soit consacrée aux activités de recherche pour l'évolution des produits existants ou pour la création de nouveaux produits. Cette enveloppe – qui pourrait s'apparenter à ce qui existe actuellement pour le Groupe de Mission Mercator Coriolis – permettrait de lancer annuellement des appels à idées ou des appels d'offres, qui seraient évalués et suivis par les comités scientifiques des pôles de données identifiés plus haut.

Il est par ailleurs proposé de traiter l'expertise liée à l'utilisation des données et produits hors des pôles de données sous forme participative, au travers d'un fonctionnement en réseau, et d'échanges de bonnes pratiques, notamment lors d'ateliers utilisateurs.

Proposition n° 6 – S'appuyer sur des moyens mutualisés à différents niveaux

La mutualisation des ressources d'un pôle de données dépend des fonctions qu'il doit assurer et de leur organisation, au sens de l'architecture physique du pôle.

Parmi les contraintes associées, figurent les volumes et la variété de données à manipuler et les moyens de stockage, d'archivage et de traitement qui y sont liés, ainsi que la nécessité pour certains jeux de données d'associer étroitement – et donc de ne pas dissocier – acquisition, traitement et qualification des observations et expertise.

Ce qui conduit nécessairement à des fonctions centralisées au niveau national (voire européen), à d'autres fonctions distribuées géographiquement (notamment dans des centres de données primaires proches de l'acquisition des observations) et à la nécessité pour les pôles de travailler avec une approche basée sur des fonctions et activités distribuées.

La mutualisation des moyens des pôles de données doit, si et quand elle est applicable, être une règle.

Elle doit notamment s'appliquer à deux niveaux :

- le niveau national, qui assure la coordination de l'ensemble des fonctions d'un pôle de données, au travers d'une structure et d'une organisation ayant vocation à assurer ces fonctions nationales. *A cet effet, il est souhaitable que chaque pôle national de données s'appuie – au moins pour la coordination des fonctions techniques et scientifiques, sur une structure propre à vocation nationale de type Unité Mixte de Service (UMS),*
- le *niveau décentralisé où, pour les fonctions et activités de gestion des données, les fédérations de laboratoires (de type OSU) doivent jouer un rôle majeur* et disposer des ressources dédiées à la vocation nationale des activités (et d'obtenir la reconnaissance qui va avec).

La distribution des fonctions des pôles de données doit être traitée au cas par cas pour chaque pôle de données, en s'assurant que les ressources (notamment humaines) contribuant à des activités de niveau national – ou européen – sont effectivement dédiées à ces activités.

Les moyens affectés aux pôles de données nationaux ne doivent pas être confondus avec les moyens – tout aussi nécessaires – affectés aux unités de recherche pour l'utilisation des données ou produits d'observation distribués par les pôles.

Proposition n° 7 – Des liens entre la recherche et l'opérationnel à définir clairement

Il est essentiel que les pôles de données en géosciences, qui ont vocation à servir les communautés de recherche, n'empiètent pas sur les mandats des opérateurs de service – que ceux-ci soient institutionnels ou marchands – et ne dupliquent pas les activités et fonctions de ceux-ci.

Ceci n'interdit pas que les données et produits d'observation distribués par les pôles de données soient utilisés hors de la sphère recherche, dans la mesure où ceux-ci sont accompagnés d'une licence d'utilisation permettant de dégager les pôles de données des responsabilités liées à leur utilisation. Cela est vrai en particulier quand les pôles de données génèrent et distribuent des produits élaborés et qualifiés qui répondent à des besoins non couverts par d'autres opérateurs.

C'est en particulier le cas d'observations « recherche » utilisées pour répondre à des obligations réglementaires, ou dans des services européens de type Copernicus.

Par contre, le *niveau de service* d'un pôle de données ne peut être compatible avec les contraintes d'opérateurs de service, en particulier dès qu'il s'agit de fournir des services *en temps réel* ou devant être qualifié par un opérateur mandaté.

Il est également nécessaire de considérer que l'expertise issue d'activités de recherche sur les données ou produits d'observation s'est constituée dans les institutions de recherche, et doit y être maintenue tant qu'elle n'est pas transférée à un opérateur spécifique.

Il est important de définir le niveau d'expertise qui doit être préservé dans les structures de recherche lors d'un transfert d'activité et de compétence vers les structures opérationnelles (opérateurs), en particulier quand cette expertise est nécessaire aux activités de recherche.

Les pôles de données, et les organismes de recherche qui y sont impliqués, doivent donc (i) identifier systématiquement les données sur lesquelles ils possèdent une expertise et qui passeront en configuration opérationnelle, (ii) s'associer avec les opérateurs potentiels afin de maintenir l'expertise nécessaire lors du transfert vers l'opérationnel et (iii) s'assurer que ce transfert ne s'accompagnera pas d'une dégradation du niveau de qualité des données au détriment des utilisateurs « recherche ».

Proposition n° 8 – Un rôle bien défini pour les structures privées

Les structures privées peuvent être associées à des pôles de données au travers :

- d'activités demandant des *compétences classiques* en systèmes d'information et en gestion de données, telles que les développements informatiques (codage ou optimisation de logiciels), l'administration de systèmes d'exploitation ou la production (traitement ou archivage de données),
- des activités exigeant une *expertise spécifique* (étalonnage/validation, contrôle qualité, documentation scientifique), souvent acquise dans une structure de recherche et transférée à la structure privée.

Les tâches standard de gestion de données qui ne peuvent être réalisées par les institutions (notamment par défaut de personnel) peuvent être confiées à des structures privées.

Par contre, lorsqu'il s'agit d'expertise directement liée à la valeur ajoutée du pôle, il est nécessaire, avant contractualisation avec une structure privée :

- de préserver une « masse critique » d'expertise interne au pôle sur les thèmes ou fonctions concernées, et d'organiser les priorités de recrutement du pôle en fonction de cette préservation,
- de demander à la structure privée un engagement sur la pérennisation de l'expertise concernée, y compris et si nécessaire avec l'aide des institutions qui externalisent ces activités.

Les pôles de données ont également vocation, au travers des données qu'il génèrent et diffusent, et de leur expertise, à servir les structures privées, notamment pour leur permettre de développer et valider des applications ou services à vocation commerciale.

Cette approche pourrait être soutenue par des structures traitant du transfert recherche-industrie, telles que les pôles de compétitivité nationaux ou les Sociétés d'Accélération de Transfert Technologique (SATT).

Proposition n° 9 – Les données des pôles doivent être utilisées pour la formation et la communication

La valorisation des données d'observation passe par leur utilisation dans le domaine de la formation secondaire ou universitaire, mais également par la génération et la diffusion d'informations destinées au *grand public* – et facilement compréhensibles par celui-ci. La qualité des données et produits d'observation offre des possibilités insuffisamment exploitées de description *pédagogique* de phénomènes ou processus – voir par exemple le site web *OMER7* qui propose un outil pédagogique utilisant les données et informations du pôle Ether.

Par ailleurs, la communication institutionnelle prend de plus en plus d'importance, notamment en cas d'événement catastrophique ou de crise. Les pôles de données et leurs experts peuvent intervenir – directement ou indirectement – dans la fourniture d'informations pouvant être communiquées au grand public lorsque survient ce type d'évènement.

De plus, les réseaux sociaux, qui prennent une importance grandissante, notamment dans les jeunes générations, peuvent démultiplier la diffusion, voire l'interprétation, de l'information d'origine.

Il est donc nécessaire que les organismes et institutions définissent quelle est la politique de communication et de diffusion de l'expertise autour des pôles de données, pour des objectifs pédagogiques mais aussi en cas d'évènements à forte portée médiatique.

Ces approches de valorisation, importantes pour la visibilité de la société civile sur les activités de recherche liées aux observations, doivent être effectivement mises en place, et doivent s'appuyer sur des actions, et des moyens associés bien identifiés. Il est notamment nécessaire d'acquérir, là où elles n'existent pas, les compétences spécifiques pour ce type d'activités.

Les activités similaires en place dans d'autres pays peuvent servir de modèles pour les pôles de données nationaux.

Propositions à caractère technique

Les propositions techniques mettent l'accent sur certaines fonctions ou processus des pôles de données pour lesquels des évolutions ou actions correctives sont nécessaires ; ces propositions ne s'appliquent pas nécessairement à tous les centres de données actuellement en place.

Proposition n° 10 – L'accès aux données : une fonction à améliorer

Les fonctions liées à l'accès aux données et produits des pôles sont évidemment cruciales pour ceux-ci : elles représentent une interface *stratégique* avec les utilisateurs des pôles, en permettant notamment d'identifier et de caractériser (par les données qu'ils demandent) ces utilisateurs et de recueillir leurs retours sur l'usage des données.

Elles qualifient également la facilité et la souplesse d'accès aux données des pôles.

De plus, ces fonctions sont étroitement liées à la découverte et à la documentation des données des pôles, ainsi qu'au support scientifique ou technique (expertise) que les pôles apportent à leurs utilisateurs.

Contrairement à certaines institutions ou structures non françaises qui ont abordé la gestion de données pour les géosciences avec une logique de point d'accès unique aux données – acquérant parfois de ce fait une position incontournable au niveau européen et international – cette fonction a été souvent mal structurée ou négligée au niveau français, par exemple en multipliant les points d'accès aux données (portails) ou en offrant des accès limités à des *initiés*.

Si des moyens d'accès aux données simples et robustes peuvent être privilégiés – en particulier quand ces données s'adressent à des communautés nationales restreintes – un effort de structuration des fonctions de présentation et d'accès aux données doit être consenti au niveau national, en particulier autour de l'approche de pôles de données *intégrateurs* proposée précédemment.

Cela suppose notamment, pour l'ensemble des pôles de données, le développement de systèmes *dynamiques* d'accès aux données (mise à jour des catalogues, informations) s'appuyant sur un réseau de plusieurs sources de données et capables de gérer des accès différenciés – par exemple pour des jeux de données faisant l'objet de licences d'utilisation.

Par ailleurs, les données et leurs documentations (*métadonnées*) doivent, en particulier si elles sont destinées à un usage ouvert (par exemple au niveau international), se conformer à des standards ou principes – qui peuvent être réglementés, par exemple dans le cas de la directive INSPIRE de l'UE – à des recommandations internationales (par exemple GEOSS ou OMM) ou à des pratiques *de facto* par certaines communautés.

L'utilisation de Digital Object Identifiers (DOI) permettrait également de mieux mesurer l'usage et l'impact des données distribuées par les pôles, tout en permettant à terme une meilleure reconnaissance de l'activité et de la contribution des producteurs des données.

Enfin, les contraintes d'identification des utilisateurs doivent – quand elles sont nécessaires – être simplifiées au maximum (par exemple en utilisant des techniques de Single Sign On), et harmonisées pour chaque pôle, voire au delà de ceux-ci...

Proposition n° 11 – L’archivage des données : une fonction cruciale à gérer au niveau national

Les volumes croissants (en instantané et en cumulé) et la diversité des observations acquises et gérées au niveau national rendent cruciale la question de leur archivage pour leur utilisation immédiate ou différée, mais également pour la sauvegarde du patrimoine qu’elles représentent.

Il faut distinguer les besoins d’archivage court/moyen terme – ou *stockage* – de ceux d’archivage long terme – ou *pérennisation*.

Toutes les données doivent être archivées à court/moyen terme, cataloguées et documentées ; un inventaire de toutes les données spatiales et *in-situ* doit être effectué et un plan à court terme (3 ans maximum) doit permettre d’éviter les données orphelines. Les pôles de données doivent coordonner l’archivage à moyen terme, qui peut être physiquement décentralisé, dans les centres de données dotés des moyens adéquats en termes de disponibilité et de sécurité.

La pérennisation des données d’observation est nécessaire, car d’une part elles constituent un patrimoine unique et non reproductible et, d’autre part, on ne peut définir précisément aujourd’hui l’intérêt qu’elles pourront présenter à long terme.

La pérennisation des données requiert des moyens spécifiques, souvent lourds, liés aux volumes et à la variété des données en question et à l’usage qui peut en être fait. Cette question, pour laquelle la priorité doit être accordée aux observations de base à caractère patrimonial, doit être instruite au niveau national.

Il est proposé qu’un plan de pérennisation des données (besoins, solutions) soit préparé dans les deux années à venir, sans ralentir les efforts engagés sur les données des missions spatiales en fin de vie ou des données in-situ « dormantes ».

La recherche d’une solution unique – basée sur une institution agissant au niveau national par délégation des autres institutions concernées et couvrant à la fois les données des missions spatiales et d’autres sources – doit être privilégiée, afin de réaliser des économies d’échelle et de moyens.

Proposition n° 12 – Des processus de développement à redéfinir

Les pôles de données appliquent, pour leurs chaînes de traitement des observations, des cycles de vie consistant en (i) une phase de mise au point de ces chaînes de traitement, (ii) une phase de consolidation correspondant notamment à la validation scientifique et techniques des chaînes, suivie (iii) d’une phase de production en routine, où les chaînes de traitement sont en place, et où l’on se place dans une logique de production systématique.

Ces cycles de vie sont parfois mal respectés, notamment pour les traitements de nouvelles missions spatiales, et ce principalement à cause de l’inadéquation des processus de développement actuellement utilisés. Les phases de mise au point et de consolidation des chaînes de traitement des données sont parfois contraintes par des choix d’organisation, en particulier de sous-traitance industrielle sur la base de spécification.

Il est proposé que les entités (laboratoires ou organismes de recherche) possédant l’expertise sur les méthodes de traitement pilotent la phase de mise au point des chaînes de traitement et co-pilotent – avec les entités en charge des systèmes de production – la phase de consolidation de ces chaînes.

Pour ce faire, une phase de définition et de maquettage des chaînes algorithmiques doit être identifiée, s’appuyant sur des moyens informatiques et d’accès aux données adaptés (bacs à sable, outils de

prototypage, environnements de développement). Durant cette phase, la plus grande souplesse de traitement et de validation doit être disponible, au travers de calculateurs et moyens de gestion des données dédiés au développement, ouverts et facilement accessibles à la demande.

L'industrialisation des chaînes de traitement ne doit pas être abordée avant la phase de consolidation de celles-ci, et doit s'appuyer sur un « durcissement » progressif des spécifications et des méthodes de développement (utilisation d'approches de type « Agile »).

Proposition n° 13 – Prendre rapidement en compte l'arrivée de nouvelles technologies

Les besoins fondamentaux des pôles de données sont (i) de disposer d'architectures opérationnelles pour les chaînes de production, qui garantissent la disponibilité de la production et la sécurité du stockage, (ii) de proposer un accès unifié à la donnée indépendamment de sa localisation, (iii) d'offrir la capacité de déployer des chaînes de traitement utilisant des données multi thématiques, et (iv) de faire facilement face à des demandes massives et temporaires de ressources (pour les retraitements par exemple).

Les nouvelles technologies de gestion et de traitement des données (par exemple le *cloud computing*) vont profondément modifier l'organisation des fonctions et ressources des pôles de données.

Elles permettront notamment, au travers d'infrastructures de type *data center*, de mutualiser les ressources d'archivage et les moyens de traitement. L'activation de certaines fonctions se fera de manière déportée à partir de terminaux virtuels (et éventuellement nomades).

Ces infrastructures, qui sont adaptées à la gestion de volumes importants de données d'une grande diversité, doivent être dimensionnées pour assurer la disponibilité requise. La mise à profit de technologies de type *interopérabilité et virtualisation (cloud communautaire)* doit permettre de rendre ces ressources disponibles à la communauté utilisatrice, indépendamment de leur localisation physique, et moyennant l'utilisation de réseaux de données adaptés pour relier les pôles et centres de données.

Ces principes de mutualisation doivent permettre une réduction substantielle des ressources allouées au Maintien en Condition Opérationnelle des infrastructures, tout en offrant la souplesse requise par les utilisations pour les géosciences.

Il est nécessaire d'analyser l'utilisation et l'impact des nouvelles technologies de gestion des données pour assurer les fonctions des pôles, définir l'architecture de leurs systèmes d'information, dimensionner les ressources nécessaires et définir un calendrier de mise en œuvre.

La réalisation ou la consolidation de ces objectifs doit se faire sans augmentation des coûts complets de possession des infrastructures, tout en améliorant la qualité de service.

Proposition de feuille de route
pour la mise en œuvre
des propositions du rapport

Objectifs

L'objectif de la feuille de route est de proposer un certain nombre *d'étapes et d'actions à court terme* liées à la mise en œuvre des propositions du groupe de réflexion.

Cette feuille de route est principalement destinée aux organismes qui auront, s'ils le souhaitent, la responsabilité de mise en œuvre des propositions du groupe de réflexion. La prise en compte urgente de certaines propositions, notamment celles liées à la stratégie européenne, constitue une des conditions d'efficacité et de succès des propositions.

Les étapes *court terme* de mise en œuvre des propositions

Les propositions du groupe visent à mettre en place, à *moyen terme*, une nouvelle organisation nationale autour de la gestion de données pour les géosciences, basée sur quatre pôles de données et une coordination scientifique et technique inter-pôles.

Cette échéance à moyen terme – *qui peut se traduire par un délai de l'ordre de cinq ans et modulable selon chaque pôle* – est nécessaire pour :

- d'une part réorganiser et/ou faire évoluer les périmètres et les différentes fonctions techniques et scientifiques des différents pôles de données pris individuellement ou considérés dans leur ensemble (notamment pour les aspects inter-pôles), et redimensionner les ressources affectées à ces fonctions ;
- d'autre part, simplifier les processus de gouvernance existant autour de ces activités, en supprimant certains niveaux de décision intermédiaires, et en les intégrant au bon niveau de visibilité et de synthèse.

Les années 2014-2015 seront cruciales pour mettre sur les rails cette nouvelle configuration nationale, et toute perte de temps et de moyens à court terme se traduira nécessairement par des retards préjudiciables à moyen terme.

Avec cette perspective d'ensemble à moyen terme, les principales échéances à court terme sont les suivantes :

A l'automne 2013

- acceptation de principe – avec éventuellement amendement – par les organismes concernés des propositions contenues dans le rapport
- accord sur les actions à engager d'ici fin 2014

Fin 2013

- la configuration de travail est en place
- les acteurs sont identifiés et leurs mandats et échéances sont définis

Mi-2014

- la proposition de stratégie européenne pour les pôles de données est définie
- les organismes décident sur les éléments de cette stratégie

fin 2014

- les rapports définissant la configuration à moyen terme des quatre pôles de données, ainsi que les étapes et moyens pour y parvenir sont rendus
- les organismes décident sur le plan d'action à moyen terme

Actions à court terme, acteurs et échéances

A. Mettre en place la configuration de travail – échéance fin 2013

La configuration de travail proposée pour cette phase de définition et construction consiste en :

- ***un comité de pilotage inter-organismes***, chargé des prises de décision de niveau institutionnel,
- ***pour chaque pôle, un comité de mise en place***, chargé de définir les caractéristiques et l'organisation du pôle en tenant compte de l'existant (utile / non utile à maintenir / subsidiaire,...), ainsi que les étapes de structuration spécifiques à ce pôle et la feuille de route associée,
- ***un groupe technique inter-pôles***, chargé d'étudier les éléments et activités techniques mutualisables entre les différents pôles,
- ***pour chaque pôle, un chargé de mission*** dont le rôle est d'analyser le périmètre du pôle et la mise en place de son organisation, et de sa politique scientifique et technique au niveau national et européen, et de faire des propositions sur ces points au *comité de mise en place du pôle en question*.,

Comme leur dénomination l'indique, les comités de mise en place et les chargés de mission sont associés à cette phase *court terme* et ne devraient pas être pérennisés.

Pour mettre en place cette configuration, il est proposé de réaliser les actions suivantes :

A1. Réunion des organismes pertinents pour approbation – et éventuellement amendement – des propositions du rapport et décision sur la feuille de route – octobre 2013

- Le CNES et le CNRS/INSU convoquent les organismes et si nécessaire des représentants des ministères concernés.

- Les représentants des organismes et institutions recueillent les avis et amendements sur les propositions.
- Ils décident de mettre en place un comité de pilotage composé de représentants des organismes et d'un petit nombre d'experts – ils définissent les grandes lignes et échéance de son mandat – ils nomment son ou ses présidents.
- Ils décident quelles institutions doivent être impliquées dans les montages de chacun des quatre pôles – cela conduit à un comité de mise en place pour chacun des pôles – ils nomment un président pour chacun de ces comités.
- Ils décident la mise en place d'un groupe technique inter-pôles destiné à proposer les éléments et activités mutualisables entre les pôles, de sa composition et nomment son président.

A2. Réunion des comités de mise en place de chacun des pôles – novembre 2013

- Chaque comité définit les éléments particuliers de sa feuille de route jusqu'à fin 2014.
- Chaque comité définit une fiche de poste *pour un chargé de mission* qui instruira durant l'année 2014 la configuration de son pôle (périmètre, organisation, échéancier, moyens, etc.).

A3. Les chargés de mission sont nommés – décembre 2013

- Les comités de mise en place choisissent leurs chargés de mission et valident leur mandat.

B. Instruction de la stratégie européenne – échéance mi-2014

B1. Le comité de pilotage (ou un sous-groupe de celui-ci) et les chargés de mission débutent le travail sur la stratégie européenne – janvier 2014

- Ils définissent les points à instruire et les échéances.

B2. La stratégie européenne est définie – juin 2014

- Un rapport est écrit par le comité de pilotage et les chargés de mission.
- Les décisions sont prises par les institutions sur cette base.

C. Configuration des 4 pôles et aspects inter-pôles – échéance fin 2014

C1. Les quatre chargés de mission remettent leurs rapports – octobre 2014

- Ils ont interagi au préalable – au moins à mi-parcours – avec leurs comités de mise en place respectifs et avec le comité de pilotage, en particulier pour les aspects inter-pôles.

C2. Le groupe technique inter-pôles remet son rapport – octobre 2014

- Il a interagi au préalable – au moins à mi-parcours – avec le comité de pilotage et plus fréquemment avec les quatre chargés de mission.

Annexe 1

Mandat du Groupe de réflexion

Toulouse, le 27 juillet 2011
DSP – 2011.15857

Objet : Mandat pour un groupe de réflexion inter organismes sur les pôles thématiques en observation de la Terre

Contexte

Le concept de pôles thématiques en Observation de la Terre a été défini au niveau national il y a plus de 15 ans. On peut se reporter au rapport de référence dit « rapport Waldteufel » datant de 1999.

Ce concept correspondait principalement au besoin de la communauté de recherche française de disposer de capacités de développement et de génération de produits dérivés des instruments d'observation de la Terre, provenant en particulier de missions définies et opérées par le CNES.

La logique « multi-capteurs » (plusieurs instruments pour dériver un même paramètre ou étudier une thématique scientifique donnée) était fortement liée aux périmètres et fonctions de ces pôles thématiques. Par ailleurs, une approche multi-organismes au niveau français (principalement autour d'activités de recherche) était privilégiée, à la fois en termes de définition, de fonctionnement et de processus de décision pour ces pôles thématiques.

A partir de ce concept, ont été mis en place les pôles Ether et Icare, ainsi que certaines fonctions applicables à la thématique « surfaces continentales ». Par ailleurs, le service SALP, sans avoir formellement suivi le processus « pôles thématiques », remplit la plupart des fonctions d'un pôle.

Chaque pôle regroupe des fournisseurs de données, des experts scientifiques et des utilisateurs de données et bénéficie de moyens mis en place par les organismes et institutions partenaires dans une logique de mutualisation. Il est conçu au niveau national pour servir la communauté scientifique autour d'une thématique donnée et stimuler l'utilisation de données de missions spatiales, mais sans que le contexte européen ou international et la vision à long terme, qui va avec, soient pris explicitement en compte.

Depuis cette définition du concept et la mise en place d'un certain nombre de « pôles thématiques » nationaux, de nombreux éléments ont évolué, tant au niveau national, qu'aux niveaux européen et international.

Au niveau national, les besoins scientifiques se sont affirmés en raison de l'accroissement des connaissances conduisant à de nouvelles activités et priorités, en particulier liées à l'exploitation des dimensions temporelles et spatiales des observations. Ces besoins exigent la continuité des données, engendrent la demande de nouveaux services (notamment pour la Terre solide ou la couleur de l'eau,...) et suscitent de nouvelles interfaces avec les centres de données (capteurs). Les moyens, notamment spatiaux, liés aux activités de recherche (moins de missions nationales en développement et augmentation du poids des missions en

exploitation) et l'organisation nationale de la recherche ont évolué. La demande sociétale s'accroît (Grenelle de l'environnement, Grenelle de la mer) et un processus d'opérationnalisation en trois phases, recherche-développement-services, en identifiant les acteurs principaux dans chacune de ces phases doit se mettre en place.

L'Europe a pris conscience des besoins communautaires – incluant ceux de la recherche – et met en place des programmes ou réglementations structurants, à travers l'Union Européenne (GMES, PCRD, directive Inspire, ...) ou des organisations intergouvernementales comme Eumetsat (observations spatiales pour la météorologie, le climat et l'océanographie, ...) et l'ESA (Sentinelles GMES et produits dérivés, Climate Change Initiative pour l'Observation de la Terre ; actions de normalisation, interopérabilité et pérennisation des données, ...). Le concept de centre de données dont les fonctions s'apparentent à celles des pôles thématiques français a été retenu pour GMES. La question de l'intégration des pôles thématiques nationaux dans un cadre européen reste ouverte. De même les Satellite Application Facilities d'EUMETSAT destinées à prendre en compte les nouveaux besoins des services météorologiques et les futures missions EUMETSAT évoluent, de nouvelles activités et services liés aux nouvelles missions (scientifiques et Sentinelles GMES) de l'ESA apparaissent.

Au niveau international (au-delà de l'Europe), des mécanismes de coordination existent (comme ceux de l'Organisation Météorologique Mondiale) ou sont en développement dans le cadre du Système de Systèmes du Groupe pour l'Observation de la Terre (GEOSS). Ils imposent de définir le positionnement national quant à la contribution aux activités internationales d'échanges de données d'observation de la Terre.

Ces nouveaux éléments de contexte poussent à réfléchir à une éventuelle évolution des périmètres et organisations des Pôles Thématiques et à s'interroger quant aux capacités européennes existantes ou à venir à satisfaire, au moins en partie, les besoins de la communauté française.

Par ailleurs lors de sa séance de juin 2010, le CPS avait souligné l'utilité et la validité des pôles thématiques. Afin de préparer leur insertion européenne, il recommandait « d'inventorier et d'analyser les structurations dans les autres pays, de travailler à une clarification des différents types d'usage, du traitement de données des missions jusqu'aux informations environnementales ». Ce dernier point nécessite de prendre en compte dans les réflexions sur l'évolution des pôles thématiques à la fois d'autres sources de données que celles des satellites, en particulier les données *in-situ* et les sorties de modèles, et des communautés utilisatrices élargies au-delà de la recherche (institutions et le cas échéant décideurs).

Enfin, lors de sa séance de juin 2011 le comité TOSCA, soutenu par le CPS, a exprimé le besoin d'une actualisation des « termes de références génériques » des Pôles Thématiques. Elle sera nécessaire pour maintenir une certaine cohérence entre les Pôles Thématiques qui montrent des dynamiques très diverses selon le domaine thématique concerné, leur mode de gouvernance et l'implication des organismes nationaux.

Afin de proposer des recommandations sur les actions à mettre en place pour répondre au mieux aux besoins des communautés nationales, dans un contexte national, européen et international complexe et évolutif, il est créé un groupe de réflexion inter-organismes sur les pôles thématiques en observation de la Terre.

Axes de réflexion et d'interrogation

Prenant en compte les éléments de contexte présentés ci avant, il est demandé que le groupe de réflexion

- procède à un état des lieux national, européen et international sur,
 - o les pôles thématiques, centres de données, moyens d'observation, etc... ;
 - o leurs fonctions, structures, moyens, utilisateurs, interactions, difficultés, perspectives ;
 - o ce qui se fait de similaire en Europe ou en international (en comparant les efforts consentis et les bénéfices pour les communautés utilisatrices) ;
 - o ...
- analyse les demandes des utilisateurs nationaux à l'horizon des 10-15 ans à venir en mettant l'accent sur celles qui sont prioritaires et ne sont pas, actuellement ou en prévision, satisfaites ;
- propose des schémas de collaboration inter-organismes autour des pôles thématiques ou leur évolution :
 - o analyse fonctionnelle, architecture générale ;
 - o proposition de termes de références génériques ;
 - o organisation : pilotage, gouvernance et décision ;
 - o ...
- propose une stratégie d'intégration européenne (éventuellement par thème) voire internationale :
 - o fonctions à conserver ou à distribuer (partenariats internationaux) ;
 - o sources de financement possibles ;
 - o impacts financier pour les organismes nationaux ;
 - o ...
- propose un plan d'action national sur l'ensemble des éléments analysés en mettant en évidence,
 - o les priorités de mise en œuvre ;
 - o les moyens associés et les acteurs concernés ;
 - o ...

Composition du groupe

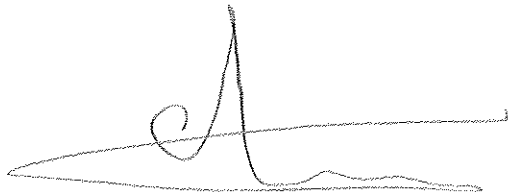
Président : Alain PODAIRE
Membres : François VIAL (INSU)
Didier ROUMIGUIERES (CNES)
Steven HOSFORD (CNES)
Jean-Pierre GLEYZES (CNES)
Gilles BERGAMETTI (TOSCA/LISA)
Nicole PAPINEAU (IPSL)
Jean Pierre WIGNERON (INRA)
Vincent Henri PEUCH (CNRM – METEO)
Philippe BERTRAND (INSU)
Michel DIAMENT (INSU) ou son représentant

Le groupe de réflexion s'appuiera - sous forme de contributions ou d'auditions - sur l'expertise existant dans les laboratoires et organismes français.

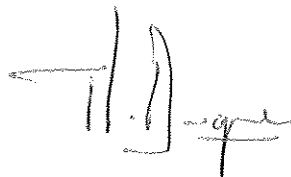
Calendrier

Il est demandé au groupe de réflexion d'organiser son travail sur un an avec un point d'étape lors du colloque de restitution proposé par le TOSCA les 21 et 22 mars 2012.

La présentation finale des conclusions du groupe sera faite au CPS d'automne 2012. Elle pourra être précédée d'une discussion préalable avec le TOSCA.



Pour l'INSU
Jean François STEPHAN
INSU/D



Pour le CNES
Thierry DUQUESNE
DSP/D

Diffusion :

CNES : DSP/DA, DSP/OT, DCT/D, DCT/DA, DCT/PS, DCT/ME, DCT/SL, DCT/PO, participants au groupe.

INSU : JM Flaud, M. Diamant, E. Verges, G. Guyot, participants au groupe
Autres participants au groupe.

Annexe 2

Personnes auditionnées par le Groupe de Réflexion

Audition « Etat des lieux » - décembre 2011

- Vincent Cassé (LMD) – Services d’observation (SO) « Atmosphère »
- Philippe Ciais (LSCE) – ICOS Atmosphère
- Cathy Clerbaux (LATMOS) – ESPRI/IPSL
- Guy Delrieu (LTHE) – SO OHM-Cévennes Vivarais
- Jérôme Gaillardet (IPG Paris) – SOERE « Bassins versants »
- Thierry Guinle (CNES) – Service altimétrie spatiale
- Yann Kerr (CESBIO) – CATD SMOS Humidité
- Pascal Kosuth (IRSTEA) – GEOSUD et Pôle « Surfaces continentales »
- Bernard Legras (LMD) - ICARE
- Denis Loustau (INRA) – ICOS Surfaces
- Helle Pedersen (Institut des Sciences de la Terre de Grenoble) – RESIF : réseau sismologique et géodésique français
- Mioara Manda (CNES) – Pôle « Formes et mouvements de la Terre »
- Sébastien Payan (LATMOS) – ETHER
- Sylvie Pouliquen (IFREMER) – Coriolis
- Nicolas Reul (IFREMER) – CATD SMOS Salinité
- Gilles Reverdin (LOCEAN) – SO « Océan »
- Hervé Roquet (Météo-France) – OSI SAF et SATMOS
- Benoit Sautour (EPOC) – SOMLIT
- Jean Virieux (Institut des Sciences de la Terre de Grenoble) – SO « Terre solide »

Audition « Besoins » - janvier 2012

Besoins des thématiques et programmes de recherche

- Pierre Brasseur (LGGE) – océanographie hauturière
- Eric Ceschia (CESBIO) – carbone et gaz à effet de serre à l’échelle locale
- Véronique Ducrocq (Météo-France) - météorologie
- Jean-Louis Dufresne (LMD) – climat
- Antoine Gremare (EPOC) – océanographie côtière
- Isabelle Manighetti (GeoAzur) – risques
- Vincent-Henri Peuch (CEPMMT) – chimie atmosphérique

Besoins des structures

- Yann Kerr – CESBIO
- Hervé Le Treut – IPSL
- Hubert Loisel, François Schmitt - LOG
- Eric Maneux – Geotransfert
- Patrick Raimbault - MIO
- Nikolai Shapiro – IPG Paris
- Pierre Soler – OMP
- Wolfram Wobrock – LaMP

Audition « Organismes » - octobre 2012 / février 2013

- BRGM : Alain Bosse, François Robida, Jean-Jacques Serrano
- CEA : Cyril Moulin
- CNES : Richard Bonneville, Olivier Marsal, Pascale Ultré-Guerard
- CNRS / INSU : Jean-François Stephan
- IFREMER : Pierre Coty, Patrick Farcy, Marie-Hélène Tusseau-Vuillemin, Patrick Vincent
- INRA : Marie Rabut, Jean-François Soussana
- IRD : Robert Arfi
- IRSTEA : Pierrick Givone, Pierre Maurel
- Météo-France : Philippe Bougeault
- MNHN : Sylvie Dufour, Sylvie Rebuffat, Jean-Dominique Wahiche

Audition « Structures privées » - février 2013

- ACRI : Jérôme Bruniquel, Odile Fanton d'Andon
- Astrium Geoinformation Services : Jean-Michel Darroy, Bruno Montfort
- CLS : Fabien Lefèvre, Gilles Larnicol
- CS : Anne Chanié, Thierry Rabaute
- HYGEO : Didier Ramon
- Noveltis : Richard Bru
- Telespazio : François Marques, Nicolas Vincent
- Thalès Alenia Space : Jean-Guy Planès

Autres auditions

- Mark Thorley – National Environmental Research Council, Royaume Uni
- Mike Daniels, Jeff Stitt – NCAR, Etats-Unis
- Frank Flechtner, Reinhard Hüttl, Herman Lürr, Ludwig Stroink – GFZ, Allemagne
-
- Nicolas Arnaud – INSU
- Xavier Briottet, Christian Martin – ONERA
- Philippe Dandin, Isabelle Donet, Bruno Piguet – Météo-France
- Laurence Fleury - OMP
- Laurence Rouil – INERIS

Annexe 3
Moyens existants
dans les pôles de données

Introduction

Une enquête approfondie – mais pas nécessairement exhaustive – a été réalisée pour les quatre futurs pôles de données. L’objectif était de cartographier l’ensemble des moyens existants autour de la gestion des données. Au final, seules les ressources humaines ont pu être analysées. Le questionnaire envoyé aux centres de données ou aux responsables de systèmes ou services d’observation *in situ* portait principalement sur les moyens attribués aux fonctions définies pour un pôle de données, à savoir :

- acquisition des données (validées, traitées après acquisition sur le terrain, satellite...) dans les centres de données,
- développement de méthodes, d’algorithmes, de codes de calcul,
- production prototype des données,
- production systématique et retraitement des données,
- contrôle qualité des données et produits,
- mise à disposition des données, de produits, ou de codes de calcul,
- mise à disposition de ressources techniques,
- support utilisateur, formation,
- archivage,
- pérennisation.

Nous avons reçu des réponses de la majeure partie des centres de données interrogés.

Quelques constats et constantes

La situation rapportée aux quatre futurs pôles de données nationaux met en évidence des différences significatives entre ceux-ci (existence de pôles et centres de données organisés ou non, maturité des réseaux d’observation, ...), avec quelques constantes, dont notamment :

- une bonne mutualisation des moyens lourds de gestion de données – chaque pôle de données s’appuie sur quelques grands centres de données,
- une forte implication des laboratoires et équipes de recherche dans les fonctions nécessitant un expertise scientifique,
- le fait que le traitement des données des Services Nationaux d’Observation *in situ* est essentiellement effectué dans les laboratoires, car il nécessite une expertise proche de la donnée ; dès qu’il y a besoin de moyens de production et de distribution plus importants, les SNO s’appuient sur les OSU ou les pôles / centres existants afin d’assurer ces fonctions.

Les ressources humaines actuelles

L’enquête conduite auprès des différents centres de données a permis d’évaluer les ressources humaines actuellement en place autour des différentes fonctions relevant des futurs pôles nationaux de données.

Notons que le nombre d'ETP (Equivalent Temps Plein) pour chaque pôle peut paraître élevé, mais il correspond au total des personnes impliquées dans chaque SNO ou centre de données, et inclut de fait des personnels des laboratoires souvent impliqués avec de faibles pourcentages de temps dans les expertises et traitements.

Gestion des données « Atmosphère »

Le tableau 1 détaille les centres de données pris en compte dans l'analyse.

Type de structure	Structures prises en compte
Pôles	ICARE et CES (Megha-Tropiques & Calipso, LOA & IPSL) Ether (OMP & IPSL)
Campagnes	AMMA, HYMEX
Centres	SEDOO (ATMO), ESPRI, OSU-PDD, SATMOS
SO	CESOA, SIRTa, NDACC, RAMSES/ICOS IAGOS, IDAF, PAES, AERONET/PHOTONS CO-PDD

Tableau 1

Le tableau 2 fournit une synthèse des ressources humaines (en ETP) pour les différentes fonctions relevant du pôle de données *Atmosphère* – incluant la gestion des données de campagnes non strictement liées à l'atmosphère (AMMA, HYMEX).

Fonctions	Pour tous types de structures		
	Total Chercheurs + ITA	Total Chercheurs	Total ITA
Fonctions techniques : production, distribution, web, archivage, etc...	45	7	38
Fonctions scientifiques : développement méthodologique, validation, expertise...	59	25	34
Total ETP	104	31	73

Tableau 2

Une première analyse, à affiner, montre que :

- les fonctions techniques, mutualisées dans quelques centres de données, regroupent un peu moins d’ETP que les fonctions scientifiques ; ce qui est logique car le nombre de Services d’Observation est important et requiert en majorité des fonctions d’expertise scientifique,
- il existe quelques centres de données à vocation nationale, tels que les CGTD ICARE et ETHER, mais aussi ceux de quelques OSU qui disposent de moyens suffisants pour assurer cette vocation,
- les fonctions nécessitant une expertise scientifique de proximité sont du ressort des laboratoires et équipes de recherche – elles correspondent à des fractions d’ETP qui finissent par faire masse une fois cumulées,
- dans les fonctions d’expertise, la proportion de chercheurs impliqués est importante : cela reflète le fait que ces tâches sont parfois uniquement assurées par des chercheurs pour pallier au manque d’ITA,
- un début de rationalisation de la production et la diffusion des données *in situ* est en cours mais il est nécessaire d’amplifier le mouvement ; il y a une demande et un besoin de regroupement des fonctions de distribution – et parfois de production – dans les centres de données importants (OSU., etc.),
- le nombre d’ETP identifiés peut paraître élevé mais est représentatif de la situation. En reprenant les données reportées dans le tableau 2, les 70 ETP ITA correspondent à environ 35 ETP pour les grands centres de données et campagnes, 15 ETP dans les laboratoires pour les fonctions d’expertise, de traitement, de contrôle qualité, et de validation liées aux missions spatiales et environ 20 ETP pour les 7 SNO (qui concernent de nombreux instruments). Ce nombre d’ETP inclut une vingtaine de CDD mis en place au démarrage de nouvelles activités, par exemple Megha-Tropiques, ou qui permettent de faire face au besoin croissant de traitement des données d’observation.

Gestion des données « Surfaces continentales »

Le tableau 3 détaille les centres de données qui ont été pris en compte.

Type de structure	Structures prises en compte
Pôle	THEIA
SO	AMMA CATCH HYBAM BVET H+ SONEL AgrHyS ACBB OSR TOURBIERES KARST OHMCV RENAG OHGE GIACIOCLIM CORAIL DRAIX ONEVU OBSERVA BVEA BVED LTC Misanthrop/GIFSI SODYC SYN APSE
Centres	OREME THETA INRA GRAVI OSU Rennes OHGE

Tableau 3

Le tableau 4 fournit une synthèse des ressources humaines pour les différentes fonctions relevant du pôle de données.

Fonctions	Pour tous types de structures		
	Total	Total Chercheurs	Total ITA
Fonctions techniques : production, distribution, web, archivage, etc...	42	10	33
Fonctions scientifiques : développement méthodologique, validation, expertise...	25	8	16
Total ETP	67	18	49

Tableau 4

Les éléments recueillis montrent que :

- le nombre de SO est important et qu'ils sont, pour un nombre significatif d'entre eux, récents ; de manière générale, une activité importante est consacrée à la validation des données recueillies, mais peu de traitements lourds sont appliqués. Ces constats expliquent que les fonctions liées à la distribution des données apparaissent peu développées. Environ 15 ETP sont identifiés à ce jour sur ces activités, parmi lesquels un faible nombre est affecté à des fonctions techniques.
- les données des SNO peuvent apparaître comme dispersées, mais un effort de mutualisation a été fait – et doit se poursuivre – pour la production et la distribution des données,
- quelques OSU regroupent également des données d'observation non intégrées dans des SNO labellisés, mais très utilisées pour des études locales,
- Le pôle Theia est en « construction », et il ne couvre pas pour l'instant la gestion des données *in situ*. Il est structuré autour de deux centres de données bien délimités, avec des expertises fortes dans quelques laboratoires clés. 20 ETP ont été identifiés dans la convention constitutive.

Gestion des données « Terre solide »

Le tableau 5 décrit les centres de données *in situ* prises en compte.

Type de structure	Structures prises en compte
SNO	GEOSCOPE, GRAVIMETRIE, INSTABILITES MAGNETISME RENAG, RENASS, VOLCAN

Tableau 5

Le tableau 6 fournit une synthèse des ressources humaines pour les différentes fonctions relevant du pôle de données – limitées aux SNO considérés.

Fonctions	Pour SNO seulement		
	Total	Total Chercheurs	Total ITA
Fonctions techniques : production, distribution, web, archivage, etc...	17	10	7
Fonctions scientifiques : développement méthodologique, validation, expertise...	12	2	11
Total ETP	29	12	17

Tableau 6

Commentaires :

- la gestion des données des SNO du domaine est bien organisée au plan national (assurée par les OSU) et reconnue et visible au niveau européen et international,
- il y a une forte demande de personnels pour produire et distribuer les données des SNO,
- le pôle « géodésie spatiale / formes et mouvements de la Terre » est en cours de construction et permettra de répondre aux besoins en données et produits dérivés dans ce domaine. Il devrait intégrer environ 30 ETP, non pris en compte dans le tableau 6.

Gestion des données « Océan »

Le tableau 7 présente les centres de données pris en compte.

Type de structure	Structures prises en compte
Pôles	SALP CORIOLIS CADTS
SO	Argo France OISO/CARAUSS/PIRATA SOMLIT MOOSE
Centres	Ifremer ROSCOFF OASU

Tableau 7

Le tableau 8 fournit une synthèse des ressources humaines pour les différentes fonctions relevant du pôle de données, qui couvre les domaines hauturier, côtier et littoral.

Fonctions	Pour tous types de structures		
	Total	Total Chercheurs	Total ITA
Fonctions techniques : production, distribution, web, archivage, etc...	67	6	61
Fonctions scientifiques : développement méthodologique, validation, expertise...	14	4	10
Total ETP hors acquisition données	81	10	71

Tableau 8

Commentaires :

- la gestion des données spatiales est organisée par variables océaniques (altimétrie, température, salinité, couleur...). Le futur pôle national devrait permettre de mettre en place une approche plus intégrée.
- chacun des centres de gestion des données spatiales s'appuie sur des expertises dans les laboratoires. La maturité des centres de données et des communautés utilisatrices explique le peu d'ETP recensés sur les fonctions scientifiques ; les activités d'expertise ont en partie été transférées dans les communautés utilisatrices des données ;
- la gestion des données *in situ* pour le domaine hauturier est bien organisée nationalement via Coriolis, qui fait l'objet d'une convention inter-organismes et qui intègre progressivement l'ensemble des données des SNO,
- la gestion des données *in situ* des domaines côtier et littoral, qui inclut les activités de SOMLIT et des stations marines se structure progressivement. Une approche nationale intégrée, impliquant l'ensemble des organismes et institutions pertinents, reste toutefois à appliquer et constituera un enjeu majeur de la mise en place du futur pôle national de données océaniques.

Annexe 4

Éléments sur les enjeux, l'état des lieux et les besoins

L'annexe 4 du rapport fournit des éléments sur l'analyse des enjeux scientifiques, de l'état des lieux et des besoins pour les compartiments atmosphère, océan et surfaces continentales.

L'atmosphère

Enjeux scientifiques

Les recherches dans le domaine atmosphérique concernent principalement la dynamique atmosphérique, la météorologie et la chimie atmosphérique. Elles incluent également, comme pour les autres compartiments, des travaux plus orientés vers l'étude des évolutions du climat. Ces thématiques concernent à la fois la troposphère et la stratosphère, chacune de ces régions atmosphériques étant subdivisée autant que nécessaire selon les problématiques abordées.

Il existe, pour des raisons évidentes, une très longue expérience en matière de sciences météorologiques (au sens prévision du « temps » à court terme) : depuis plus d'un siècle, des observations systématiques de température, humidité, vitesse de vent, pression... sont réalisées dans de nombreuses stations dédiées de par le monde. Des vecteurs particuliers (ballons sondes par exemple) ou d'opportunité (bateaux de commerce, avions de ligne, etc.) ont été instrumentés pour documenter également les zones les moins accessibles (océan ouvert, couches supérieures de l'atmosphère, etc.). Plus récemment, les observations satellitaires sont venues très fortement renforcer ce dispositif d'observation en fournissant une vision spatiale plus étendue.

En parallèle, ont été organisés l'archivage, la mise à disposition et le partage de ces données par la communauté des météorologistes. Les observations ont été fiabilisées et pérennisées, notamment celles nécessitant des moyens récurrents lourds comme les observations satellitaires, des modèles de circulation atmosphérique à différentes échelles ont été développés, l'assimilation des données a été mise en œuvre et des services opérationnels ont été mis en place, avec un emboîtement de tâches bien définies entre les niveaux internationaux, nationaux, régionaux et locaux. Par ailleurs, la dimension opérationnelle a été identifiée et dimensionnée en fonction des besoins très forts d'utilisateurs pourtant très variés (agriculteurs, transporteurs, militaires, etc.). La dimension recherche reste très présente, tant pour le développement de nouveaux outils d'observation que pour la réalisation de campagnes d'observation intensives permettant de mieux cerner les phénomènes, en particulier de petite échelle, et de pouvoir ainsi les représenter de façon plus pertinente dans les modèles de prévision du temps. On voit donc au travers de ce bref survol que la météorologie s'est déjà confrontée à beaucoup des enjeux et difficultés que rencontrent d'autres domaines des géosciences, et que les solutions mises en œuvre fonctionnent plutôt bien, au moins pour les objectifs propres à ce domaine. La météorologie constitue donc un exemple de très grand intérêt pour beaucoup de questions qui se posent à d'autres domaines de recherche ou d'application : densité et pérennité des observations, partage des données, liens recherche/opérationnel, cadre de coopération internationale, services aval...

Les études climatiques en lien avec l'atmosphère relèvent pour partie des mêmes démarches. Il s'agit là encore de comprendre et de prévoir l'évolution, non plus du temps mais du climat. Les échelles de temps sont différentes car bien plus longues. Pour cela, et pour ce qui concerne la composante atmosphérique, il est évident que les changements de la composition de l'atmosphère en espèces traces (CO_2 , CH_4 , N_2O , O_3 , aérosols, etc.) doivent être suivis au travers d'observations, d'où la mise en place aux niveaux national et international de réseaux d'observation denses et inscrits dans la durée (RAMCES, NDAC, AERONET, etc.).

Les conséquences possibles des modifications de la composition de l'atmosphère sur des variables clés comme la température et les précipitations, ou sur les caractéristiques d'objets à fort pouvoir de rétroaction comme les nuages, sont également étudiées, notamment sur une échelle décennale, grâce aux données archivées et aux observations satellitaires. Enfin, des modèles tridimensionnels, incluant les autres compartiments du système Terre, visent à évaluer les conséquences à long terme de ces évolutions de la composition de l'atmosphère. L'importance accordée à ces travaux est bien plus récente que pour ceux conduits en météorologie et donc, logiquement, l'organisation et la structuration

est moins achevée, notamment en ce qui concerne l'identification des champs d'application et la façon de gérer les éventuels services associés. On notera néanmoins de – trop ? – nombreuses initiatives ou instances visant à organiser et fédérer au niveau international ces recherches.

En matière de chimie atmosphérique, les enjeux ont considérablement évolué au cours des dernières années. Ainsi, l'étude de l'évolution de l'ozone stratosphérique a connu dans les années 1980-1990 une phase de recherches intenses sur l'identification des causes et mécanismes de sa destruction. Ces recherches, associant tous types d'outils (ballons, avions, satellites, simulations expérimentales, modèles), ont conduit à beaucoup d'avancées et ont permis de décrire assez précisément ces mécanismes. Associés à l'adoption du protocole de Montréal limitant la fabrication des CFC responsables de la destruction de l'ozone stratosphérique, ces résultats concluants ont contribué à faire évoluer l'étude de la stratosphère, d'une part vers un simple suivi des niveaux d'ozone pour évaluer la reconstitution de l'ozone stratosphérique, et d'autre part vers des études davantage orientées sur l'impact climatique d'un refroidissement de la stratosphère.

Aujourd'hui, en raison d'études sanitaires inquiétantes et de textes législatifs contraignants, les problématiques en chimie atmosphérique sont surtout orientées vers l'étude de la pollution photo-oxydante et particulaire. Selon les échelles concernées, différentes couches de l'atmosphère présentent un intérêt plus ou moins important : à l'échelle globale, où les problématiques de pollution photo-oxydante vont rejoindre pour partie des problématiques climatiques (l'ozone troposphérique est, par son importance, le troisième gaz à effet de serre), c'est essentiellement sur la troposphère libre que se porte l'attention. A l'échelle locale ou régionale, là où les concentrations de polluants et de particules sont telles qu'elles peuvent avoir des effets nocifs sur la santé humaine ou sur les écosystèmes, la priorité concerne la couche limite, là où les êtres vivants se trouvent.

Organisation

Compte tenu de ces enjeux divers, beaucoup d'acteurs interviennent (agences en charge de la qualité de l'air, associations de surveillance, réseaux internationaux, etc.) pour assurer la collecte d'observations, la gestion des bases de données (sujet très sensible) et la prévision (donc la mise en œuvre d'outils numériques). Si, au niveau national, beaucoup d'efforts d'organisation et de partage de tâches ont été faits (voir le système PREV'AIR), la structuration européenne, malgré des actions dédiées à la qualité de l'air dans le programme CAFE (Clean Air For Europe) et les projets GEMS et MACC dans le cadre GMES/Copernicus concernant notamment les bases de données, reste à définir et stabiliser.

Sur le plan de l'organisation et de façon générale, la communauté nationale travaillant sur les sciences atmosphériques, tant sur la troposphère que sur la stratosphère, est une communauté assez fortement structurée : elle est pour l'essentiel regroupée au sein de laboratoires de taille souvent importante, eux-mêmes acteurs ou parties prenantes de structures communautaires ou fédératives comme les OSU et autres fédérations de recherche, dont l'OMP et l'IPSL sont les plus représentatifs. Elle dispose de deux pôles thématiques (ICARE et ETHER), qui ont plus plusieurs années d'existence. Cette forte structuration a permis le rassemblement de chercheurs et d'ITA en nombre significatif, souvent de métiers et de compétences variés et complémentaires. Même si des améliorations sont souhaitables, il existe donc indubitablement, dans les laboratoires « atmosphériques » au sens large, une capacité importante de traitement et de gestion de données d'observation, y compris satellitaire. De même, beaucoup de ces laboratoires ont la capacité de produire des algorithmes qui vont permettre aux chaînes de traitement de données satellitaires de générer les produits ad hoc. Cette forte structuration est également associée à une longue culture d'utilisation de données de campagnes *in situ*, de données de services d'observation ou d'observations satellitaires.

De fait, aujourd'hui, les structures de recherches (laboratoires, OSU, Fédération de recherche, Labex, etc.) ont l'expérience et des personnels compétents pour traiter jusqu'à un niveau assez amont les différentes données, y compris les données satellitaires.

Ce constat, bien plus net aujourd'hui qu'il ne l'était il y a une dizaine d'années lors de la publication du rapport dit « Waldteufel », induit quelques évolutions dans la façon dont les utilisateurs de données atmosphériques expriment leur attente vis-à-vis des pôles communautaires de traitement et gestion de données. En particulier, il semblerait que la demande de produits élaborés (niveaux 3 ou 4) par un centre dédié soit moins forte qu'elle ne l'était il y a une dizaine d'années, les laboratoires se sentant davantage capables aujourd'hui de réaliser ces produits par eux-mêmes et « sur mesure ». A l'inverse, les demandes de production fiable, fortement validée, et d'archivage très propre des données de niveau 1, sont plus fortement exprimées que par le passé. On veut être sûr de travailler sur des données indiscutables, sans avoir la nécessité de les vérifier systématiquement. En fait, les laboratoires (et les chercheurs) sont aujourd'hui davantage en difficulté devant les traitements de base à faire subir aux flux de données importants issus des observations systématiques, réalisées tant au sol que depuis l'espace, que dans l'analyse géophysique des observations.

Une autre spécificité des activités conduites en matière de recherche atmosphérique est liée à la diversité des champs traités et à l'existence d'acteurs divers, souvent partenaires, en matière de production et de gestion de données. La recherche météorologique est un bon exemple de ces partenariats divers, tant au niveau national qu'européen ou international. Lorsqu'on rajoute à la dimension « recherche » les aspects opérationnels, la production et l'organisation des jeux de données deviennent alors d'une grande complexité, y compris examinées du seul point de vue de la recherche. On retrouve, à un degré moindre, le même type de complexité mais avec des acteurs différents (ou pour certains dans des rôles différents) dans l'étude de la qualité de l'air. En fait, on pourrait dire de façon sans doute un peu caricaturale, mais néanmoins pas totalement fautive, que les domaines de la recherche et de l'opérationnel ont, en sciences atmosphériques, fait de tels progrès chacun dans leur domaine qu'ils ont beaucoup à échanger en termes d'outils et de données. Cependant, les organisations permettant de faire cela simplement restent encore et en dehors de la météorologie, si ce n'est à inventer, au moins à optimiser au cas par cas.

Enfin, le contexte international est également assez spécifique. Il existe des initiatives visant à fédérer un certain nombre d'observations, voire à produire des données élaborées ou à fournir des services liés à l'usage de ces données. Si l'organisation et les rôles sont bien définis au niveau de la météorologie entre le CEPMMT, EUMESAT et les services météorologiques nationaux, c'est loin d'être le cas dans les autres domaines d'intérêt atmosphérique. On note bien le positionnement de l'Agence Européenne de l'Environnement sur les bases de données, le rôle structurant des projets GMES/Copernicus GEMS et MACC pour la production d'information géophysique, mais une vraie organisation générale et intégrée reste à définir et surtout à faire adopter.

L'océan

Le domaine océanique peut être subdivisé en trois sous-domaines : hauturier, côtier et littoral. Les stratégies d'observation des variables océaniques – qu'elles soient physiques ou biogéochimiques – présentent des caractéristiques différentes dans ces trois sous-domaines :

- pour les **systèmes hauturiers**, l'objectif est de décrire l'état tridimensionnel de l'océan de la grande à la méso-échelle et de comprendre les phénomènes et processus qui s'y déroulent : l'échantillonnage spatial et temporel nécessaire s'appuie sur des observations spatiales – essentiellement pour la surface – et *in situ* – pour la surface et la dimension verticale de l'océan ;
- pour les **systèmes côtiers**, qui surmontent les plates-formes et les pentes continentales, l'objectif est identique à celui des systèmes hauturiers, mais à des échelles plus fines, et avec des apports provenant du littoral et des surfaces continentales (bassins versants). Les performances actuelles des observations spatiales, notamment en termes de résolution spatiale et de répétitivité, rendent leur utilisation plus limitée dans ces systèmes ; c'est donc principalement sur des observations *in situ* – avec leurs limitations en termes d'échantillonnage – que repose aujourd'hui la description des systèmes côtiers.
- les **systèmes littoraux** incluent la frange littorale terrestre (dunes, plages, falaises, rives d'estuaires, deltas...), la zone intertidale pour les mers à marée, et une zone infratidale (fonds inférieurs à 20 m) dont la morphodynamique est en continuité avec celle des zones littorales émergées. Ils sont le lieu d'interaction entre les systèmes continentaux et océaniques, avec des contraintes importantes liées aux activités humaines : leur description, qui doit notamment prendre en compte ces interactions, doit intégrer l'ensemble des variables océaniques et continentales pertinentes. La réalisation exhaustive de cette description, notamment à l'échelle du territoire national, ne peut se concevoir que sur des missions spécifiques, hors recherche, comme l'inventaire LITTO 3D actuellement réalisé par le SHOM par imagerie LIDAR bathymétrique aéroporté. Néanmoins un tel inventaire n'est pas adapté à l'étude de processus : cela explique que, pour la recherche, les études littorales sont réalisées au travers de zones ateliers sur lesquelles sont concentrées les observations.

L'état des lieux dans le domaine hauturier

Le domaine hauturier a tiré un double bénéfice de la disponibilité depuis 20 – et parfois plus de 30 – ans d'observations de la surface océanique depuis l'espace et du développement, puis de la mise en œuvre de l'océanographie opérationnelle (en France par Mercator Océan, puis au niveau européen au travers du programme Copernicus), basée en partie sur la disponibilité de ces observations spatiales. Les systèmes d'observation *in situ*, présents avant l'arrivée des missions spatiales, mais avec un échantillonnage spatial souvent limité, ont accompagné cette montée en puissance, et atteignent maintenant un niveau comparable à celui des systèmes spatiaux, que ce soit au travers de réseaux internationaux, où la participation française est décisive (tels Argo ou SSS), ou de services d'observation avec une vocation « recherche » plus affirmée (tels que Pirata ou Memo).

L'important investissement national en océanographie opérationnelle a permis de structurer un ensemble de centres de données, où la prise en compte de la contrainte « temps réel » est dimensionnante :

- les centres de données spatiales SALP pour l'altimétrie (en place depuis près de 20 ans) et le SAF « Ocean and Sea Ice » (mis en place depuis plus de 10 ans au travers d'EUMETSAT)

pour la température de surface, les caractéristiques des glaces de mer, et les flux océan-atmosphère ;

- le centre de données *in situ* Coriolis, qui rassemble les données de réseaux permanents et pérennes, collectées depuis des navires ou des systèmes autonomes, fixes ou dérivants, tel que le système Argo qui acquiert des mesures sur la dimension verticale de l'océan.

Ces centres de données, qui sont bien établis dans le paysage national et européen, fournissent, au delà des utilisateurs opérationnels, un service reconnu aux communautés de recherche. Par contre, comme ils fonctionnent en mode « projet » ou au travers de conventions à durée limitée, leur pérennité n'est pas garantie.

D'autres sources de données existent à côté de ces centres de données bien visibles et reconnus. Elles incluent notamment des capacités autour de données de couleur de l'océan et d'optique marine, pour lesquelles une approche structurée et pérenne au niveau national (et européen) reste à mettre en place. Le GIS COOC pourrait préfigurer cette approche, mais son statut en limite pour le moment la portée. Enfin, l'existence de nouvelles missions permettant d'estimer la salinité de surface des océans (notamment SMOS, qui dispose d'un centre de données, le CATDS) doit conduire, après qualification, à intégrer cette variable dans les futurs centres de données océaniques.

L'état des lieux dans le domaine côtier

Comme précisé précédemment, les observations du domaine côtier sont – pour le moment – principalement collectées par des systèmes *in situ*, à l'exception des imageurs spatiaux possédant une résolution spatiale adéquate pour caractériser la température de surface, l'état de surface (vague et vent) ou la couleur de l'eau.

La collecte d'observations *in situ* au niveau national repose notamment sur des réseaux ou des infrastructures d'observation plus localisées, dont certains ont le label « Service d'observation ».

Les observations côtières *in situ* sont réalisées à partir de plusieurs types de plates-formes : des navires océanographiques, des bouées automatisées fixes, des navires d'opportunités (ferrys, navires de pêches), des radars HF basés à terre, des planeurs sous-marins (gliders).

Le SOERE MOOSE (Mediterranean Ocean Observing System on Environment) coordonne l'ensemble des données récoltées pour le domaine nord-ouest méditerranéen (zone Ligure et Golfe du Lion). Ses données sont gérées par Coriolis. Il est à noter que MOOSE est à la fois un dispositif côtier (notamment pour la marge du Golfe du Lion), mais également hauturier (zone Ligure et grands fonds). D'autres dispositifs côtiers n'entrent pas encore dans un cadre régional ou national coordonné (par exemple Ferrybox ou Recopesca). Leurs données, qui comportent souvent des observations spécifiques (notamment biologiques et biogéochimiques), sont actuellement gérées sur des bases *ad hoc* ; leurs données physico-chimiques « classiques » sont transférées à Coriolis.

L'état des lieux dans le domaine littoral

La recherche et les dispositifs d'observation associés au domaine littoral concernent plusieurs types de thématiques, a priori très déconnectées, mais qui tendent aujourd'hui se réunir dans des programmes intégrés portant sur des zones ateliers (un estuaire, une lagune, un littoral sableux ou rocheux, ...), ou même sur des ensembles régionaux plus larges dépassant éventuellement le strict cadre littoral (par exemple à la dimension d'aires marines protégées, de parc marins, etc.).

De telles évolutions d'approches sont bien sûr liées aux échanges croissants entre la communauté scientifique, les élus et les gestionnaires environnementaux.

Parmi les thématiques principales, on peut notamment citer les questions relatives à :

- l'évolution du trait de côte et à la vulnérabilité aux submersions qui s'adressent aux hydro- et morpho-dynamiciens ;
- la stabilité des habitats et de la diversité biologique des zones littorales, qui s'adressent aux biologistes et écologues, mais aussi aux physiciens et aux chimistes ;
- la santé des écosystèmes et la santé humaine qui s'adressent aux écotoxicologues, mais aussi aux physiciens et aux chimistes.

Pour toutes ces thématiques, l'observation spatiale, embrassant à la fois les domaines continentaux, littoraux et côtiers, est amenée à jouer un rôle croissant (dynamique du trait de côte à l'échelle régionale ou au delà, cartographie des habitats, suivis de blooms toxiques, suivi de panaches de contaminants, ...). Quelques zones tests existent déjà, comme celles sélectionnées au travers du programme Kalideos Littoral du CNES.

La particularité des données *in situ* littorales est leur très grande diversité. Elles incluent des données physiques et physico-chimiques (comme pour les domaines hauturier et côtier), mais aussi des données biologiques, chimiques et biogéochimiques plus ou moins complexes et composites (descriptions taxonomiques et éthologiques, composition de populations et de peuplements, données génomiques, descriptions d'habitats, concentrations et spéciations de contaminants, niveaux et types de contaminations, données isotopiques...), ainsi que des données topologiques géo-référencées des zones émergées ou immergées (altimétrie aérienne, bathymétrie, nature des substrats, couverts végétaux...). Bien entendu, le schéma se complexifie encore considérablement dès que l'on intègre les données sociales et économiques (cartographie des habitats urbains littoraux, implantations d'activités agricoles, industrielles ou touristiques, etc.).

En raison de leur complexité, les données littorales sont aujourd'hui, pour l'essentiel, gérées localement et de manière non coordonnée.

Il y a toutefois plusieurs exceptions :

- le Service (national) d'Observation du Milieu LITtoral (SOMLIT), labellisé par l'INSU, qui réunit 8 stations marines sur des activités d'observation basse fréquence normalisée à points fixes (13 paramètres normalisés, plus des paramètres locaux exploratoires), et qui a développé une base de données nationale spécifique ;
- les données taxonomiques des stations marines françaises sont désormais intégrées dans deux bases de données nationales qui ont été développées spécifiquement dans le cadre des actions RESOMAR (REseau des Stations et Observatoires MARins), l'une pour le compartiment benthique (BENTHOS), l'autre pour le compartiment pélagique (PELAGOS) ;
- les données topo-bathymétriques recueillies par LIDAR aéroporté dans le cadre du projet LITTO 3D, et les données altimétriques recueillies par LIDAR aéroporté dans le cadre du projet CLAREC.

Un enjeu important pour l'évolution des dispositifs d'observation littoraux est de fédérer en réseau les plates-formes automatisées haute fréquence développées dans de nombreuses stations, et souvent déployées aux points SOMLIT (ou proches). Il s'agira notamment de normaliser et d'inter-calibrer certains paramètres, et de centraliser la gestion des données, tout en respectant les procédures propres liées aux spécificités locales et/ou à certains financeurs.

Pour les données d'observation littorales, les enjeux principaux seront donc :

1. d'en assurer la visibilité et l'accessibilité ;
2. d'en assurer l'interopérabilité, au niveau des données ou des métadonnées, en ayant à l'esprit les différents types d'utilisateurs possibles, d'une part ceux de la communauté scientifique « recherche », d'autre part les gestionnaires, usagers, et décideurs politiques ;
3. de renvoyer à des centres de compétences, notamment pour l'accès à des données composites et complexes comme en biologie, écologie ou éco-toxicologie.

Les surfaces continentales

Diversité et complexité

Les surfaces et interfaces continentales (SIC), sont un ensemble complexe et hétérogène de par :

- la diversité des thématiques qui sont concernées : agriculture et forêt, ressources en eau, cycle du carbone et de l'eau, biodiversité, aménagement du territoire, littoral, glaces continentales, urbain...
- le continuum d'échelle concerné par ces thématiques (plante, parcelle, territoire, global, etc.) ;
- la variété des acteurs et des degrés d'articulation avec la gestion opérationnelle des SIC et le *reporting* réglementaire, qui implique laboratoires de recherche, décideurs locaux (collectivités territoriales), nationaux et internationaux, opérateurs publics, bureaux d'étude liés à l'aménagement du territoire et à la gestion des ressources, etc.

Cette diversité se retrouve largement au niveau de la collecte des observations, et de l'opération des infrastructures qui y sont associées.

Les observations physiques et biogéochimiques sont collectées principalement par des systèmes spatiaux et par des réseaux *in situ*. La mise en relation et/ou le développement de synergies dans l'utilisation des observations provenant de ces deux types d'infrastructures, qui ont bien souvent peu d'interactions car demandées et opérées par des communautés souvent disjointes, reste largement à développer.

Notons que l'utilisation de capteurs aéroportés reste assez ponctuelle et limitée à des campagnes expérimentales s'étalant le plus souvent sur quelques jours d'observations sur une – ou plus rarement plusieurs – années pour des validations méthodologiques (ALPILLES RESEDA sur l'assimilation multi-capteur haute résolution, CAROLS pour la mission SMOS, etc.).

Les données spatiales

Les observations issues de capteurs opérés depuis l'espace couvrent une large gamme de résolutions spatiales : de quelques dizaines de centimètres pour les capteurs imageurs THR (Très Haute Résolution), utilisables pour des applications en biologie et écologie, sur l'urbain, etc. à des résolutions de l'ordre de 2 à 30 km pour des missions avec des applications à l'échelle globale en hydrologie et météorologie (imageurs Météosat, diffusiomètres ERS et autres, SMOS, etc.).

Pour ces missions spatiales, la répétitivité temporelle des observations est quasi inversement proportionnelle à la résolution spatiale (répétitivité d'un an pour une couverture exhaustive avec des données THR, et de quelques heures ou quelques minutes pour les données basse résolution acquises en géostationnaire comme Météosat).

Pour servir l'ambitieux programme Copernicus de l'UE, de nouveaux capteurs avec des capacités très innovantes sont attendus, à l'exemple de Sentinel 2 (ESA) à couverture globale, haute répétitivité (~ 2-5 jours) et haute résolution (~10 m).

Une autre contrainte importante pour les SIC est que la continuité des observations n'est pas forcément assurée. Cela peut venir du fait que le capteur spatial en fin de vie n'est pas remplacé ou de la complexité pour produire une série continue homogène à partir de capteurs qui dérivent dans le temps, avec des technologies et performances non homogènes (cf. la série de données de NOAA/AVHRR, qui fait encore l'objet d'analyses très pointues pour produire une série homogène sur plus de trente ans).

La communauté « recherche et développement » française impliquée dans l'utilisation de la donnée spatiale apparaît très hétérogène et parcellisée (plus de 400 laboratoires d'universités et d'établissements, plus de 50 écoles doctorales). Il y a aussi une grande hétérogénéité de maturité dans la maîtrise de la donnée spatiale et de son utilisation. Ainsi, aux côtés d'une vingtaine de laboratoires ou groupements spécialisés, une multitude de petites structures travaillent sur des applications thématiques. Le plus souvent, ces petites structures n'ont pas de compétence particulière sur la qualification ou les prétraitements appliqués aux images ou données d'observation qui, par suite, sont souvent analysées et interprétées sans traitement spécifique avant utilisation (par exemple sans corrections des effets atmosphériques).

Les modes et politiques de diffusion des données spatiales sont aussi très divers. Les données à moyenne et basse résolution sont généralement gratuites et disponibles sur des portails web gérés par des opérateurs nationaux ou internationaux, ou des universités (Météosat auprès d'EUMETSAT, MODIS auprès de la NASA, données diffusiomètre auprès de l'Université Technologique de Vienne, etc.).

Les données d'imagerie à haute et très résolution spatiale sont souvent payantes, ce qui limite leur accès pour les communautés de recherche. Pour faciliter leur utilisation, le CNES a développé plusieurs dispositifs destinés au développement des applications, dont le programme ISIS (données SPOT à prix recherche) ou les sites pilotes Kalideos (où des données acquises en continu sont mises à disposition gratuitement pour les organismes de recherche). Ces dispositifs restent cependant relativement méconnus des communautés utilisatrices, ou sont sous-utilisés par celles-ci.

Les réseaux de données *in situ*

Les données collectées par les systèmes *in situ* proviennent de nombreux réseaux d'observation avec des structures très variées.

Elles peuvent être acquises au travers de services ou systèmes labellisés (Services d'Observation – SO, label INSU ; Systèmes d'Observation et d'Expérimentation pour la Recherche en Environnement – SOERE, label AllEnvi ; etc.), qui assurent un suivi par des observations *in situ* régulières et pérennes sur des sites dont l'extension spatiale peut être très variable : depuis la dimension régionale pour l'OSR (région Midi Pyrénées) et OM-CV (Région Cévennes Vivarais) à une dimension continentale pour HYBAM dans les grands bassins fluviaux tropicaux.

Elles proviennent aussi de réseaux et dispositifs gérés par des organismes et services dont le mandat de couverture nationale est établi. Les domaines d'applications associés sont ici encore très variés et concernent par exemple, la surveillance sanitaire des forêts française (DSF, Département de la santé des forêts), la gestion d'un système d'information sur les sols de France (Groupement d'Intérêt Scientifique Sol, GIS Sol), la gestion d'une [base nationale des données pour l'hydrométrie et l'hydrologie](#) (Banque HYDRO), etc.

Enfin, les observations *in situ* proviennent également d'une multitude de structures de nature et de taille très diverses (équipes au sein d'organismes de recherche, structures régionales dépendant des ministères (ONF, DRAF, etc.), associations, GIP, Parcs naturels, etc.), qui conduisent des applications à l'échelle locale (espaces ou parcs naturels, réserve biologique, etc.), voire régionale (ressource forestière et agricole, etc.). Ces structures, dont l'inventaire exhaustif serait bien difficile à réaliser, sont réparties sur l'ensemble du territoire français, et gèrent des réseaux ou programmes d'observation sur une gamme de thématiques très vaste (écologie, urbanisation, risques naturels, biodiversité, gestion et préservation des ressources naturelles, etc.). En général, ces structures utilisent peu les données issues de l'observation depuis l'espace, qui nécessitent des informations (où obtenir des données? quel capteur choisir ? ...) et des compétences spécifiques (corrections géométriques pour superposer les images, atmosphériques, etc. nécessaires pour intégrer les données dans les SIG, conduire des inter-comparaisons qualitatives ou quantitatives, etc.).

Analyse et spécificités

Il y a probablement une relation de cause à effet entre la diversité des questions et approches scientifiques, ainsi que des structures de la communauté SIC, et son déficit d'utilisation de la donnée spatiale (acquise par satellite).

Ainsi, une quantité importante de données satellitaires (en particulier des séries historiques de 20 ou 30 ans) a été acquise mais reste largement sous-utilisée, du fait de problèmes liés aux prétraitements et la qualité des données, à leur diffusion et facilité d'accès, à leur documentation, au manque de logiciels de traitement adaptés, etc.

Le rapport l'Académie des Sciences sur la Recherche Spatiale (Puget, 2010) constate ainsi que « le problème de la chaîne de l'information (de la physique de mesure vers la fourniture de produits/services et leur utilisation dans les applications ou les processus de décision) est incomplètement posé et donc mal traité au niveau national » et que « alors que la demande sociétale est particulièrement forte dans cette thématique autour de la problématique du changement global et des interactions entre les changements d'origine anthropique (aménagement du territoire, adaptation des systèmes de culture, etc.) et ceux d'origine naturelle, le développement d'approches opérationnelles semble piétiner ».

Il existe une forte attente au sein de la communauté SIC pour la fourniture d'une information et pour une animation, au niveau national, sur les séries de données spatiales disponibles, les logiciels de traitement adaptés, les méthodes appropriées dans des champs d'applications thématiques très divers, les appels d'offres et les formations relevant du domaine spatial, etc.

Cette attente se justifie car, comme nous l'avons décrit, l'analyse des observations (spatiales et *in situ*) est très souvent conduite par des structures petites et dispersées qui n'ont ni les compétences, ni les moyens, pour réaliser des prétraitements appropriés sur les observations. Des images « prêtes à l'emploi » doivent être fournies au risque de freiner fortement l'utilisation des données par des communautés intéressées ou d'obtenir des interprétations biaisées.

Par ailleurs, pour de nombreuses applications, il est souvent indispensable de disposer de séries multi-temporelles interannuelles : pour détecter des changements (liés aux aléas climatiques ou aux pratiques culturelles par exemple), pour conduire une classification de l'occupation des surfaces (4 ou 5 images sont alors nécessaires à des dates bien choisies selon les types de paysages), pour un suivi dans un contexte de changement global (des séries multi-temporelles couvrant des décennies sont alors nécessaires). Cette attente pour des séries multi-temporelles longues révèle avec acuité la nécessité de disposer de séries d'images corrigées qui peuvent être analysées, inter comparées, ou fusionnées, sur l'ensemble d'une période d'étude.

De plus, même parmi les entités ou structures disposant de moyens importants existant au niveau national, les compétences restent dispersées. Ainsi, très peu de laboratoires disposent des compétences pour aborder l'ensemble d'une problématique qui va, par exemple, de l'analyse des données d'observation de la Terre sur un large domaine de longueur d'ondes (du visible aux micro-ondes) jusqu'à leur assimilation dans les modèles de fonctionnement, que cela se situe à l'échelle du paysage (site de l'OSR pour le CESBIO par exemple) ou du continent (assimilation dans les modèles de climat au LSCE ou à Météo France).

Une meilleure structuration de la communauté SIC au niveau national faciliterait donc l'utilisation des données issues des observations spatiales pour une large gamme d'applications et par des structures de nature très diverses. Une telle structuration est également indispensable pour répondre à des besoins, qui se placent à un niveau davantage stratégique, sur (i) la continuité des observations (acquises à partir de satellites et de réseaux *in situ*) et (ii) la concertation entre organismes pour assurer l'accompagnement et la valorisation des missions satellitaires depuis l'amont jusqu'à une utilisation opérationnelle des données.

Le Pôle Thématique Surfaces Continentales (PTSC)

Le programme d'activités POSTEL (Pôle d'Observation des Surfaces continentales par TELédétection mis en place par le GIP Medias-France) a constitué, de 2002 à 2008, une première tentative pour fournir à la communauté des produits bio-géophysiques utilisables pour étudier le rôle des surfaces continentales dans les cycles de l'eau, de l'énergie et du carbone. C'était la première fois qu'une structure spécifique répondait aux besoins d'une partie de la communauté SIC. Cependant, POSTEL se concentrait essentiellement sur des « échelles globales » et n'avait pas vocation à structurer et animer la communauté SIC. Suite à la dissolution de Medias-France, la question de la continuité de POSTEL s'est posée au niveau national. Le CIO-E a mandaté l'INSU pour engager des réflexions sur le contenu et l'organisation d'un futur Pôle Thématique Surfaces Continentales, avec l'implication de l'ensemble des acteurs nationaux de la recherche.

Une large concertation et un travail collaboratif ont permis de converger sur l'élaboration d'un document technique (précisant les objectifs du Pôle, ses utilisateurs, sa structure, ses produits et services, sa gouvernance), d'une version 0 de convention inter-organismes, et d'un programme d'activités pour 2012-2014. Il était ainsi proposé de mettre en place d'ici 2012 une structure scientifique et technique, dite Pôle Thématique Surfaces Continentales (PTSC), ayant pour vocation de contribuer à répondre aux besoins de la communauté scientifique nationale en matière de données, de produits, de méthodes et de formation liés à l'observation depuis l'espace des surfaces continentales. La finalité de ce Pôle est ainsi de fournir à la communauté nationale utilisatrice, au travers d'une structure collaborative multi-organismes, des produits et services répondant aux objectifs principaux du mandat du CIO-E et aux attentes de la communauté que nous avons précisées ci-dessus.

La convention PTSC a été signée en décembre 2012 par les partenaires du PTSC, à savoir le CEA, le CIRAD, le CNES, le CNRS, l'IGN, l'INRA, l'IRD, l'IRSTEA et Météo-France.

Architecture du PTSC

Une structure en réseau a été privilégié pour le PTSC, car elle semble le mieux à même de favoriser les synergies autour des données spatiales au sein d'une communauté scientifique caractérisée par de fortes parcellisation et hétérogénéité, dans ses champs d'applications thématiques, ses échelles d'analyse, ses domaines disciplinaires, ses acteurs et partenariats.

Cette structure repose sur deux composantes:

- un Centre de Gestion et Traitement des Données (CGTD).
Le CGTD est bâti autour de deux composantes localisées à Toulouse (CNES, IGN, ONERA) et Montpellier (partenaires impliqués dans l'EQUIPEX GEOSUD).
- un réseau de Centres d'Expertise Scientifique (CES), thématiques et /ou régionaux, constitué de laboratoires ou groupes de laboratoires développant des méthodes innovantes et des produits liés au spatial.

Ces CES, qui font toute l'originalité de la structure proposée, ont potentiellement une double vocation « thématique » et « régionale » :

- « vocation thématique » : ils mènent des travaux de recherche et développent des méthodes innovantes basées sur les données satellitaires sur des thématiques « surfaces continentales » (agriculture, forêt, urbain, littoral, montagne, écosystèmes, sociétés, approches territoriales, échanges surface/atmosphère, approches globales, etc.). Ils peuvent contribuer à mettre en réseau et fédérer les acteurs scientifiques au niveau national autour de ces thématiques.
- « vocation régionale » : ils contribuent à animer et structurer dans les Régions les échanges au sein de la communauté scientifique, et entre cette communauté et celle des acteurs publics et privés de la gestion des territoires. Ils gèrent des sites

observatoires (systèmes d'observation rassemblant données *in situ* et spatiales) mobilisables pour le calage et la validation des produits du Pôle.

Les CES participent, sur une base de volontariat, aux missions du Pôle. Ils sont autonomes et libres de déterminer le poids relatif accordé aux dimensions « thématique » et « régionale ». Les CES peuvent ainsi développer des chaînes de traitement spécialisées, ainsi que des produits et services adaptés aux problématiques régionales.

Les CES régionaux ont pour but de réunir institutions, chercheurs, et compétences, en s'appuyant sur les sites expérimentaux qui peuvent faire partie de systèmes d'observation nationaux. A côté de ces centres, il existe de nombreux laboratoires et équipes détenteurs de compétences reconnues dont le réseau aura besoin pour couvrir l'étendue du champ scientifique du Pôle. Cette « cartographie » de structures, ressources et compétences est appelée à évoluer sous l'effet des dynamiques que le PTSC contribuera à impulser.

Gouvernance

La structure de gouvernance du PTSC s'inspire fortement de celles qui sont mises en place pour les Pôles ICARE et ETHER, et comprend un Comité Directeur, un Comité scientifique et utilisateurs présidé par le responsable scientifique du Pôle, un bureau exécutif et un responsable technique.

Plan d'activité

Les principales activités du CGTD pour la période 2012-2014 concernent en particulier :

- le développement d'un portail permettant l'accès aux produits du PTSC,
- les procédures d'acquisition de couvertures nationales annuelles à partir de capteurs spatiaux haute résolution et le développement d'une infrastructure de traitement et de diffusion des données (dans le cadre du projet Equipex GEOSUD),
- la construction d'un Centre de Production de données multi-capteurs acquises par des instruments optiques haute résolution spatiale et temporelle (au sein du CNES).