



## TABLE DES MATIERES

Préambule	4
1 Introduction	4
2 Défis Scientifiques	5
2.1 Continuum et interfaces terre-mer	5
2.1.1 Observation et modélisation des impacts des changements globaux	5
2.1.2 Vers une climatologie de référence du milieu côtier Français	6
2.1.3 Variation du niveau de la mer	7
2.1.4 Gestion des ressources halieutiques, énergétiques et minérales :	7
2.2 Comprendre la zone critique	8
2.2.1 Explorer l'architecture dynamique de la zone critique	8
2.2.2 Découvrir les processus qui façonnent la Zone critique et sa transformation	8
2.2.3 Prévoir l'évolution des services offerts par la zone critique dans le futur	9
2.3 Territoires habités, agriculture, ville et santé	9
2.3.1 Villes et santé	10
2.3.2 L'exposome urbain	11
2.3.3 Agriculture, sécurité alimentaire et climat	11
2.4 Le système climatique et ses interactions	12
2.4.1 Interactions nuages - vapeur d'eau - aérosols	13
2.4.2 Interactions climat et surfaces continentales	13
2.4.3 Attribution des changements observés	13
2.4.4 Déterminer les flux aux interfaces	14
2.4.5 Bilans évaporation/précipitations des océans	15
2.5 Aléas et impacts sur la société	15
2.5.1 Aléas sismiques et volcaniques	15
2.5.2 Intensification des cycles hydrométéorologiques	16
2.5.3 Ressources minérales et énergétiques	17
3 Identification des verrous communs	17
3.1 Les principes FAIR	18
3.2 Traitements, calculs et services	19
3.2.1 Analyse de données	19
3.2.2 Méthodes mathématiques et Intelligence Artificielle	20
3.2.3 Fouille de données	20
4 Défis méthodologiques	21
4.1 Incertitudes et séparation des signaux	21

---

4.2	Assimilation et combinaison de données multi-temporelles et multi-disciplinaires	21
4.3	Sciences participatives	21
5	Besoins en formations et communication	22
5.1	Formation interne des acteurs de data terra :	22
5.2	Formation de la communauté scientifique :	22
5.2.1	De la récolte à la FAIRisation de la donnée	22
5.2.2	Apprentissage de l'utilisation d'ensembles de données d'observation et de modélisation	23
5.2.3	Apprentissage de l'utilisation des services (VRE) à la demande	23
5.3	Communications autour de notre objet d'étude :	23

---

## PREAMBULE

---

La stratégie scientifique de Data Terra englobe les enjeux scientifiques propres à chaque pôle thématique ainsi que les enjeux transverses. La motivation de ce document est de développer plusieurs défis transverses prioritaires aux interfaces des pôles thématiques pour lesquels les nouveaux services intégrés offerts par l'IR Data Terra sont indispensables pour surmonter les obstacles scientifiques et technologiques d'accès à la gamme complète des sources de données, à leur extraction et à leur combinaison pour développer des produits de données synthétisées de haute qualité. Il n'a pas vocation à décrire l'ensemble des défis auxquels les communautés scientifiques doivent répondre ; les prospectives scientifiques des organismes de recherche (comme celles de l'INSU, du CNES, de l'IRD) les décrivent de façon plus exhaustive.

L'infrastructure numérique développée dans le cadre du projet PIA3 ESR-EquipEx+ GAIA Data permettra de lever des verrous techniques d'intégration de données, issues des observations et de modélisation à différentes échelles, et de les adosser avec des calculateurs haute performance en s'appuyant sur le développement d'un réseau à très haut débit.

## 1 INTRODUCTION

---

La société et notre Planète font face à des changements majeurs, d'une rapidité sans précédent. La pression anthropique impose de mieux appréhender de multiples enjeux impactant nos sociétés : changement climatique, pollution, risques environnementaux et sanitaires, ressources, énergie, transition numérique... Ces enjeux appellent à un développement indispensable de connaissances sur le système Terre. Celui-ci est un système complexe composé de milieux physiques et vivants dans lesquels l'Homme occupe une place significative : ces milieux sont caractérisés par des processus opérant sur un très large continuum d'échelles de temps et d'espace et qui, souvent, interagissent entre eux. Pouvoir prévoir les évolutions et les événements extrêmes qui affectent le système Terre et leurs impacts sociétaux nécessite de connaître son histoire et de comprendre son fonctionnement. Cela passe traditionnellement par des travaux scientifiques basés sur l'analyse intégrée et croisée de nombreuses données d'expériences réalisées en laboratoire, d'observations sur terre, en mer, aéroportées ou depuis l'espace et issues de modélisations.

Les complexités des systèmes et processus étudiés d'une part et les immenses progrès dans la diversité, les résolutions et précisions des données d'autre part, font que les travaux dans une thématique ou sur un compartiment spécifique du système Terre, nécessitent de plus en plus d'intégrer et de prendre en compte des informations ou données provenant de nombreuses thématiques. Il est devenu indispensable de mettre en œuvre des approches pluri- ou interdisciplinaires qui nécessitent un accès aisé aux données qualifiées des autres thématiques mais aussi à des produits (données transformées) avec des incertitudes bien connues, utilisables facilement par des non spécialistes de la thématique considérée, comme illustré ci-après, par plusieurs exemples de ces défis transverses.

Depuis quelques années, la communauté scientifique bénéficie de nombreuses avancées et ruptures technologiques dans les dispositifs d'acquisition de mesures satellitaires et in-situ qui, en parallèle, ont conduit à un afflux sans précédent de données. C'est le cas, par exemple, avec les

développements basés sur des fibres optiques pour la mesure des déformations ou le passage de la mesure altimétrique au nadir à l'altimétrie à large fauchée. On peut encore mentionner comme exemple que l'évolution des précisions des mesures satellitaires (par altimétrie radar) des niveaux des lacs et rivières couplées à des mesures sols a permis d'analyser et de reconstituer des chroniques hydrologiques dans des régions faiblement renseignées (zones tropicales) et de mieux comprendre et prévoir l'évolution et impacts des processus hydroclimatiques. La communauté scientifique bénéficie également, ou va bénéficier dans un avenir proche, de la numérisation (2D et 3D) intensive des archives : écrits, enregistrements analogiques du passé mais aussi échantillons physiques (par exemple en paléontologie). Enfin, des données provenant des réseaux de la science participative sont désormais de plus en plus intégrées dans les études.

Les diverses communautés scientifiques étudiant le système Terre sont donc confrontées au même changement de paradigme sur la façon d'analyser les données et d'en extraire des connaissances. Le rythme d'augmentation actuel et prévisible du volume de données numériques ne peut être absorbé par les infrastructures numériques publiques ou privées existantes et crée de nouveaux défis. Il faut assurer non seulement la qualification des données au plus près des producteurs et leur FAIRisation<sup>1</sup>, mais aussi leurs traitements et analyses. Les nouvelles données génèrent de nouvelles méthodes dédiées d'analyse, notamment pour la séparation des sources, il en est de même pour les aspects liés à leur augmentation massive.

Ce nouveau paysage appelle aussi au développement de nouvelles techniques, voire de nouveaux métiers comme les *data scientists*<sup>2</sup>, et des formations spécifiques à incorporer dans les cursus d'enseignements traditionnels.

## 2 DEFIS SCIENTIFIQUES

---

### 2.1 CONTINUUM ET INTERFACES TERRE-MER

Le continuum Terre – Mer peut être défini par le périmètre géographique qui relie les bassins versants à la marge continentale à travers le réseau hydrographique. Il concentre un ensemble d'interfaces et de gradients environnementaux naturels très variés, générant une très forte hétérogénéité à différentes échelles spatio-temporelles. Cette mosaïque d'environnements est parmi les plus productives, et les plus peuplées ; la zone côtière est une source de biodiversité. L'attrait exercé par les zones côtières est une des causes de l'amplification du mécanisme de littoralisation qui menace les socio-écosystèmes à l'interface continent-océan. Mais c'est aussi le lieu d'enjeux de ressources très variées : alimentation, énergie, matériaux.

#### 2.1.1 OBSERVATION ET MODELISATION DES IMPACTS DES CHANGEMENTS GLOBAUX

Il est nécessaire d'évaluer les impacts des changements globaux sur la vulnérabilité du continuum Terre - Mer, sur les fonctions aux différentes interfaces (deltas, zones côtières, estuaires, etc.) et sur les paysages continentaux et marins. Les évolutions attendues sont notamment une réduction des flux d'eau et de sédiment, une augmentation du niveau marin relatif et des processus d'érosion du littoral. Toutefois, il y a encore beaucoup d'incertitudes sur l'évolution du continuum, eu égard aux

---

<sup>1</sup> L'objectif des principes FAIR est de favoriser la découverte, l'accès, l'interopérabilité et la réutilisation des données partagées

<sup>2</sup> Chargé de la gestion, de l'analyse et de l'exploitation des données massives

interfaces multiples impliquées. Alors que la tendance à considérer ces systèmes comme des entités distinctes reste dominante, le fonctionnement et les trajectoires d'évolution du continuum Terre - Mer face au changement global nécessitent une approche intégrée. Bien que les zones côtières et les bassins versants aient fait/font l'objet de nombreuses études et d'observations, ces suivis demeurent toutefois très ciblés et limités à un type d'environnement. C'est notamment le cas de nombreux suivis réalisés par des services d'observation de recherche, le plus souvent dans le cadre de Services Nationaux d'Observation (SNO), et de suivis réglementaires (exemple dans le cadre de la Directive Cadre Stratégie pour le Milieu Marin, DCSMM). L'enjeu est le développement d'une approche intégrée qui considère le continuum dans sa globalité et de réduire les incertitudes liées à la difficulté d'identifier et de modéliser les effets indirects et rétroactions liés aux activités anthropiques et au changement climatique, en comparaison de leurs impacts directs.

**Verrous: Pouvoir accéder à l'ensemble des données (non seulement celles produites par les laboratoires et les SNOs, mais aussi par les gestionnaires) sur un continuum donné pour favoriser l'approche intégrée ; besoin d'outils pour compiler, mettre à disposition, combiner toutes les sources d'observations (in situ/satellites/modèles) ; mise à disposition d'outils statistiques pour extraire tendances/variabilité ou révéler des événements, voire des shifts ; besoin d'interaction avec la société civile et les gestionnaires et de transferts de connaissances et d'échanges sur les changements de pratiques et les observations.**

### 2.1.2 VERS UNE CLIMATOLOGIE DE REFERENCE DU MILIEU COTIER FRANÇAIS

S'il existe des climatologies de référence, approuvées par la communauté scientifique internationale, des paramètres physico-chimiques essentiels des océans profonds, la situation est beaucoup moins mature en milieu côtier et littoral. Cela tient notamment aux limitations de l'observation satellitaire très près des côtes, à la diversité et à la disparité des mesures in-situ (mesures à basse et à haute fréquence, à des immersions et/ou des stades de la marée différents, diversité des capteurs, problème des zones découvrantes, etc.), et de l'absence d'outils de modélisation opérationnels assimilant correctement les données d'observation dans toute cette diversité. Ainsi, il n'existe pas de réponse unique et faisant référence, à des questions simples et d'actualité telles que "comment l'océan côtier se réchauffe-t-il le long de nos côtes ?". Ce constat vaut pour des paramètres comme la salinité, l'oxygène dissous, la turbidité, la chlorophylle, les états de mer etc. Un tel atlas est également essentiel pour réaliser des projections, des modèles climatiques, des scénarios d'évolution pluri-décennaux de l'océan côtier et de ses socio-écosystèmes. La réalisation d'un *atlas climatologique de l'océan côtier* le long des côtes françaises, nécessitera un vaste chantier méthodologique et de traitement de données, et devrait faire l'objet d'un développement conjoint entre les IR Data Terra et ILICO<sup>3</sup>.

**Verrous<sup>4</sup> : Besoin d'accéder à une combinaison de l'ensemble des données acquises à l'interface continent-océan (in situ, sat) ; développer des algorithmes de traitement des observations satellites à l'interface continent/océan et d'assimilation de données et des produits de visualisation statistique et d'identification d'anomalie.**

<sup>3</sup> Infrastructure de recherche littorale et côtière: [www.ir-ilico.fr](http://www.ir-ilico.fr)

<sup>4</sup> Ces verrous correspondent aux besoins scientifiques et/ou technologiques d'accès aux données qui constituent des blocages pour l'avancée des connaissances du défi considéré

### 2.1.3 VARIATION DU NIVEAU DE LA MER

La surface de la mer se situe à l'interface de plusieurs composantes du système Terre (océans, atmosphère et continents). Ainsi, le niveau de la mer est d'emblée une grandeur au carrefour des disciplines et la connaissance précise de ses variations nécessite une approche interdisciplinaire. Le suivi et la modélisation du niveau de la mer passent par la définition d'un cadre de référence précis, indispensable pour identifier et décrire correctement les contributions, les séparer de l'observation qui les intègre toutes (climatiques, météorologiques, astronomiques, rebond post-glaciaire, tectonique), les modéliser et les combiner pour étudier leur résultante, par exemple pour évaluer les risques de submersion.

Ainsi les risques de submersion marine, qui sont l'une des conséquences de l'élévation globale du niveau de la mer, dépendent également des effets locaux de la terre solide (subsidence) dus par exemple à l'exploitation des nappes phréatiques dans les zones urbanisées du littoral (et plus généralement au remodelage de ces zones littorales par l'homme), ou à la présence de failles sous-marines tsunamigéniques.

La prédiction des risques de submersion aux petites échelles, qui affectent notamment les populations habitant sur le littoral, implique donc de mieux tenir compte des spécificités locales et de données pluridisciplinaires. Les ressources en eau douce, utilisées pour l'alimentation en eau potable des villes côtières dont la population augmente, sont limitées, en quantité et qualité, et menacées par l'accroissement de la demande et l'augmentation du niveau des mers. L'évaluation de la ressource disponible et de son exploitation durable actuelle et future nécessite aussi des données et des études multidisciplinaires.

**Verrous : Besoin d'accéder aux données in situ (océan/ géosphère/ géodésie/ hydrologie/ tectonique), à des sorties de modèles (hydrologiques, surcharge océanique, rebond post-glaciaire et cycle sismique, ...) sur des zones géographiques à différentes échelles ; Amélioration de la précision de la hauteur d'eau des réseaux hydrologiques.**

### 2.1.4 GESTION DES RESSOURCES HALIEUTIQUES, ENERGETIQUES ET MINERALES :

La préservation des ressources halieutiques repose sur la nécessité d'assurer la gestion de la ressource, voire sa restauration, dans une approche de type développement durable. Cette approche se traduit pour la communauté scientifique par des défis spécifiques. Le premier concerne la préservation des écosystèmes qui passe en particulier par la mise en place de zones de gestion et la création d'aires marines protégées (AMP) dans l'océan ouvert. Les délimitations des AMP doivent inclure non seulement la variabilité spatiale observée mais aussi celle attendue dans les prochaines décennies, induite par le changement climatique. Cela implique d'accéder aux échelles spatiales fines des données (satellite et in situ) et des scénarios climatiques. Un second enjeu concerne les quatre grands systèmes d'upwelling (Eastern Boundary Upwelling Systems ; EBUS) : ces zones sont des régions biologiquement très productives couvrant moins de 1% de la surface des océans, mais fournissant jusqu'à 20% des pêches de capture mondiales. Il y a besoin d'une meilleure compréhension des liens entre la dynamique des upwellings, les écosystèmes marins et la chimie atmosphérique pour répondre à l'impératif de gérer durablement les ressources halieutiques.

L'océan, et en particulier le littoral, est aussi un lieu privilégié où des dispositifs de production d'énergie décarbonée peuvent être mis en place (champs éoliens, hydroliennes, énergie marémotrice, utilisation de la houle...). En milieu hauturier, les zones hydrothermales sur les dorsales sont aussi des lieux potentiels de ressources énergétiques. De surcroît, le sous-sol marin est riche de

ressources minérales (minerais métalliques, terres rares, granulats, ...) indispensables à la mise en œuvre de la transition énergétique. Là aussi, pour l'exploration, pour une exploitation durable comme pour l'estimation des impacts environnementaux, des approches pluridisciplinaires sont indispensables (voir section 2.5.3).

**Verrous : Besoin d'accéder aux données in situ (océan/atmosphère), à des produits couleur de l'eau sur des zones géographiques méso-échelles et aux sorties de modèles ; Nécessité d'une approche intégrative prenant en compte l'ensemble des processus des interactions terre/atmosphère/océan à toutes les échelles (alliant observations in situ et satellite, et modélisation couplée)**

## 2.2 COMPRENDRE LA ZONE CRITIQUE

L'initiative scientifique sur les zones critiques vise à promouvoir le travail de différentes disciplines scientifiques des géosciences et des biosciences sur le même objet et à développer une compréhension intégrée de la partie habitable de la planète en tant que système.

Si le contrôle du bilan d'énergie à la surface a permis de fiabiliser les prédictions de réchauffement climatique au cours des dernières décennies, la prise en compte explicite des autres flux reste encore fragile et partielle. La description des sols, de leur hétérogénéité et des interactions avec le compartiment souterrain est encore très fragmentaire ; la nature et la quantité de matière, gaz, émis par différentes sources en fonction de leur localisation sont documentées de façon incomplète, avec une précision et une représentativité difficile à évaluer. La prise en compte des interactions entre les composantes biotiques et abiotiques reste à renforcer pour une meilleure représentation du fonctionnement du système Terre. Une des grandes questions scientifiques de la zone critique est « comment modéliser le fonctionnement intégré de la zone critique et prévoir sa réponse à des contraintes environnementales au niveau local comme au niveau global » (Gaillardet et al., 2018<sup>5</sup>). L'étude et la compréhension de la zone critique posent notamment les défis suivants.

### 2.2.1 EXPLORER L'ARCHITECTURE DYNAMIQUE DE LA ZONE CRITIQUE

Il s'agit de mieux décrire et comprendre l'organisation structurale, chimique et biologique de la zone critique. Ceci implique de caractériser l'extension de la zone critique et son évolution dynamique (propriétés hydrodynamiques et chimiques des sols ainsi que leurs profondeurs), de comprendre le rôle de ses différentes interfaces et leur connectivité, de savoir quantifier l'impact de l'hétérogénéité spatiale et de l'intermittence sur les flux et les temps de résidence de l'eau et des matières associées. Il faut aussi pouvoir décrire et prendre en compte les concentrations de micro-organismes et leur rôle dans la dynamique de la zone critique.

**Verrous : disponibilité et accessibilité de cartographies 3D des propriétés (hydrodynamiques, chimiques, biologiques, etc.) du sol et du sous-sol en qualité et résolution suffisante.**

### 2.2.2 DECOUVRIR LES PROCESSUS QUI FAÇONNENT LA ZONE CRITIQUE ET SA TRANSFORMATION

Les processus qui façonnent la zone critique agissent à des échelles spatiales et temporelles allant de la seconde au million d'années, de l'échelle locale à l'échelle globale. Comprendre ces processus

<sup>5</sup> Gaillardet, J., I. Braud, F. Hankard, et al. « OZCAR: The French Network of Critical Zone Observatories ». *Vadose Zone Journal* 17, n° 1 (2018): 0. <https://doi.org/10.2136/vzj2018.04.0067>.

nécessite tout d'abord de savoir quantifier les bilans et flux d'eau, d'énergie, de carbone, de sédiments dans la zone critique à différentes échelles d'espace et de temps. Il faut aussi comprendre les cycles biogéochimiques qui orchestrent son évolution, les interactions biotiques et abiotiques et identifier comment la petite échelle influence la grande et réciproquement, où et quand (hot spots, hot moments) ces processus sont actifs et donc documenter le rôle de l'hétérogénéité. Beaucoup d'études scientifiques en lien avec la zone critique reposent sur le développement, l'alimentation, ou l'utilisation d'inventaires des émissions qui recensent la nature et la quantité de matière, gaz, émis par différentes sources en fonction de leur localisation. Mais la précision associée au calcul des émissions est parfois difficile à évaluer, celles notamment des données de base étant mal connues.

**Verrous: disposer de séries de données de long terme pour développer/calibrer/évaluer les modèles. Pouvoir décrire les hétérogénéités des processus et quantifier les incertitudes liées : développer la bancarisation des données à très haute fréquence spatiale et temporelle dans la zone critique, incluant les sciences citoyennes (en collaboration avec l'IR OZCAR<sup>6</sup> et les Zones Ateliers); disposer d'outils pour l'acquisition, l'exploitation et la mise à disposition de ces futurs lacs de données<sup>7</sup>.**

### 2.2.3 PREVOIR L'ÉVOLUTION DES SERVICES OFFERTS PAR LA ZONE CRITIQUE DANS LE FUTUR

Il faut pouvoir prévoir les réponses de la zone critique aux perturbations locales naturelles et/ou anthropiques dans un contexte de changement global et fournir des éléments d'aide à la décision aux politiques. L'ensemble des données acquises doit pouvoir être assimilé dans des modélisations intégrées de la zone critique, prenant en compte tous les compartiments, notamment le compartiment biologique et les rétroactions entre l'homme et son environnement, en allant vers des réanalyses de la zone critique. Ces outils nous aideront à comprendre les rétroactions entre les compartiments de la zone critique, l'océan et l'atmosphère et permettront de contribuer aux grands défis sociétaux posés par les changements globaux (risques naturels, alimentation, milieux de vie). La modélisation et la prévision des flux dans la zone critique sont complexes et très dépendantes de la résolution et de la précision des données disponibles. Certaines variables jouent un rôle majeur mais sont encore mal décrites aux précisions nécessaires (intensité des pluies, aérosols, profondeur racinaire, etc.). Un enjeu pour tous les compartiments est certes d'avoir des séries de données mais aussi d'identifier des gaps de données ou des redondances : l'inventaire des données via le portail Data Terra devrait permettre de visualiser les sites et d'améliorer, voire d'optimiser, les stratégies des observatoires.

**Verrous : Disponibilité et accessibilité des données en qualité et résolution suffisantes pour le forçage des modèles numériques et leur évaluation dans tous les compartiments. Disposer d'outils d'assimilation et d'interpolation des données ; disposer de scénarios climatiques et d'usage des sols.**

## 2.3 TERRITOIRES HABITÉS, AGRICULTURE, VILLE ET SANTÉ

Ce chapitre concerne la partie habitée de notre planète, incluant les zones urbaines et les territoires agricoles qui les nourrissent. Ces deux types de territoires sont très liés et présentent bien sûr de

---

<sup>6</sup> Observatoires de la Zone Critique : Applications et Recherche

<sup>7</sup> Lac de données : ensemble très volumineux de données provenant de sources multiples et hétérogènes.

fortes interactions. Le lien entre les espaces et activités agricoles et le climat et le cycle de l'eau est évident, mais les territoires urbains y sont eux aussi de plus en plus sensibles avec le changement climatique et la consommation de ressources sans cesse croissante des villes.

### 2.3.1 VILLES ET SANTE

La communauté scientifique doit se regrouper de façon très pluridisciplinaire pour étudier les systèmes urbains pour constituer une communauté des "sciences urbaines" qui est encore très mal structurée de l'échelle nationale à l'échelle internationale<sup>8</sup>. Cela doit permettre les avancées nécessaires pour améliorer la compréhension et la quantification des impacts liés à l'urbanisation à plusieurs échelles spatio-temporelles. Cela comprend non seulement les grands impacts environnementaux de l'urbanisation sur l'air, l'eau, les sols, la biodiversité, l'énergie et le climat régional avec des impacts sanitaires (pollution de l'air, sécurité alimentaire, nouveaux contaminants ...), mais aussi la question des impacts de la ville sur les ressources (eau, air, sol, alimentation...) et les risques, dont les risques émergents dans les environnements urbains.

Il y a de nombreux exemples où le changement climatique a été incriminé comme une cause de maladies ou de décès : mortalité liée à la chaleur ; maladies liées aux aérosols ou aux insectes, qui se propagent dans des zones géographiques devenues favorables (Lyme, malaria, dengue, méningite, ...). D'autres conséquences, essentiellement négatives, sont attendues sur la santé et le bien-être de l'homme, et les écosystèmes en général, en lien avec le changement climatique (accroissement des températures ; augmentation des événements climatiques extrêmes : vagues de chaleur, inondations, sécheresses, tempêtes).

Pour répondre à ces questionnements, les études doivent être menées à différentes échelles spatio-temporelles (de l'individu à l'écosystème, de la parcelle à la région) et en combinant plusieurs champs disciplinaires pour comprendre la fonctionnalité des écosystèmes urbains. Cela passe par l'alimentation de bases de données diverses (inventaires d'émissions, qualité de l'air, température de surface, occupation des sols, biodiversité, épidémiologie, ...) multi-paramètres et à des échelles spatiales et temporelles variées. Les moyens mis en œuvre à l'avenir doivent bénéficier de l'évolution de nouvelles instrumentations et d'études à des échelles très fines et à des pas de temps courts. Par exemple, l'évolution de la métrologie (screening, empreintes chimiques, amélioration des capacités de détection, échantillonneurs et capteurs mobiles individuels et en réseau, capteurs satellitaires à très hautes résolutions spatiales et temporelles) permettra une meilleure perception et quantification des niveaux de contamination et d'exposition. De plus, ces suivis devront permettre d'évaluer les bénéfices ou désavantages de nouveaux aménagements afin de proposer des indicateurs susceptibles d'aider la prise de décision dans les villes pour promouvoir l'adaptation des hommes, des animaux et des plantes au changement climatique.

Les bilans globaux des impacts sont nécessaires en prenant en compte l'extérieur de la ville (de sa périphérie jusqu'à la région) et les zones de transition (ex gestion des sols urbains et péri-urbains, dépendance alimentaire de la ville à son arrière-pays). Une approche interdisciplinaire doit être favorisée, associant par exemple génie civil et urbanisme, à l'économie, la géographie, la sociologie, la santé, l'écologie, les sciences du climat et de l'atmosphère...

---

<sup>8</sup> Ghezzehei, T., Jones, J., Gholz, H.L., Vereecken, H., Van Looy, K., 2018. Steering operational synergies in terrestrial observation networks: opportunity for advancing Earth system dynamics modelling. *Earth Syst. Dynam.*, 9(2): 593-609. DOI:10.5194/esd-9-593-2018

**Verrous : Besoin de bases de données diverses (inventaires d'émissions, qualité de l'air, température de surface, occupation des sols, biodiversité, épidémiologie, ...) à des échelles spatiales et temporelles variées. Besoin de développement de modélisations numériques (et leur validation) intégrant différentes échelles de processus (du territoire dans sa globalité jusqu'à l'individu, c'est-à-dire son échelle spatiale et les déterminants de son exposition).**

### 2.3.2 L'EXPOSOME URBAIN

Le concept d'exposome correspond à la totalité des expositions à des facteurs environnementaux (c'est-à-dire non génétiques) que subit un organisme humain de sa conception à sa fin de vie en passant par le développement in utero, complétant l'effet du génome. Il illustre assez bien les difficultés pour tenir compte de manière intégrée des différentes expositions du citadin et de leurs sources (air, eau, alimentation, médicaments, radiation, ...) et la nécessité de disposer d'outils « intelligents » pour croiser et interroger de multiples bases de données, souvent hétérogènes et construites avec des finalités ou des origines très différentes. La végétation urbaine, l'hydrologie urbaine, le microclimat urbain ou la pollution urbaine sont déjà étudiés, mais une structuration de l'observation sur ces milieux urbains reste à faire. De plus, l'étude de la ville à l'échelle du quartier, de la rue, du logement va générer des données massives afin de permettre les études couplées observation / modélisation.

**Verrous : Besoin de développement de modélisations numériques intégrant des échelles très fines. Besoin de gérer et d'analyser des données massives d'origine et de nature diverses.**

### 2.3.3 AGRICULTURE, SECURITE ALIMENTAIRE ET CLIMAT

L'agriculture mondiale affronte des défis multiples<sup>9</sup>. Elle doit augmenter sa production de 70% d'ici 2050 pour accompagner l'accroissement de plus de 30% de la population et la modification des préférences alimentaires. Simultanément, l'agriculture doit faire face aux changements climatiques, aux événements extrêmes, à l'évolution de la demande et aux pressions pour réduire son impact sur les sols, l'eau, la biodiversité, le climat et pour contribuer à l'atténuation du changement climatique.

*« Si les systèmes alimentaires et agricoles continuent à évoluer en ligne avec les tendances actuelles, il est désormais avéré que l'avenir sera caractérisé par une insécurité alimentaire persistante et une croissance économique non durable<sup>10</sup> »*

Répondre à cet enjeu nécessite de mobiliser la quasi-totalité des disciplines des sciences dites dures (biologie, chimie, agronomie, écologie, hydrologie, météorologie, climatologie, ...) mais également les sciences humaines et sociales. Produire plus, produire mieux, réduire la malnutrition, s'adapter au changement climatique tout en contribuant à son atténuation, est autant affaire d'organisation politique et sociale, d'économie que de technique agricole. Le défi que pose l'évolution de l'agriculture ne peut être relevé qu'en s'appuyant sur les avancées dans d'autres domaines mentionnés dans ce document (« Comprendre la zone critique », « Climat et cycle de l'eau »).

<sup>9</sup> FAO, 2017 : L'avenir de l'alimentation et de l'agriculture : Tendances et défis - [www.fao.org/publications/fofa/fr](http://www.fao.org/publications/fofa/fr)

<sup>10</sup> FAO. 2018. *L'avenir de l'alimentation et de l'agriculture – Parcours alternatifs d'ici à 2050*. Résumé. Rome. 64 pp. <http://www.fao.org/3/CA1553FR/ca1553fr.pdf>

La grille d'analyse ci-dessous repose sur la dimension géographique, de la parcelle agricole et du bassin versant au monde entier. Cependant, la dimension temporelle est présente à toutes les échelles spatiales, pour décrire les processus, la gestion saisonnière ou les scénarios.

Echelle	Enjeux
<b>Parcelle</b>	Fournir à l'agriculteur les outils d'aide à la décision qui lui permettront, par exemple, de choisir la culture à semer en fonction de la demande, de ses ressources (sol, eau, main d'œuvre, équipement), de la prévision météorologique saisonnière, et d'optimiser les intrants (eau, fertilisants, pesticides)
<b>Bassin versant</b>	Allocation de l'eau pour l'irrigation, alerte phytosanitaire, et sur un plan plus stratégique création de réserves d'eau, mesures de réduction de la dégradation des sols et de stockage de carbone
<b>Etats</b>	Anticipation des crises et politiques assurantielles, mesures générales d'adaptation au changement climatique et d'atténuation (ex : initiative 4 pour 1000 <sup>11,12</sup> ), statistiques régionales et suivi des marchés mondiaux, suivi de la consommation de terres agricoles.
<b>International</b> <sup>13</sup>	Niveaux de production et sécurité alimentaire globale, anticipation et gestion des crises, stabilité des prix et lutte contre la spéculation, régulation du commerce, échange d'informations, de connaissances, de bonnes pratiques et d'outils.

**Verrous : besoin de stockage, de gestion et de traitement de volume de données très conséquents, incluant la dimension temporelle : séries longues de données, pluriannuelles, associées à des pas de temps courts nécessaires pour capturer les événements déterminants pour l'agriculture, et à haute résolution à l'échelle de la parcelle agricole. Traitements proches de la donnée pour éviter des transferts de données longs et des stockages multiples de l'information. Besoin d'accès aux données in-situ nécessaires à l'étalonnage des modèles, à l'apprentissage des algorithmes d'IA, et à leur validation.**

## 2.4 LE SYSTEME CLIMATIQUE ET SES INTERACTIONS

La compréhension du système climatique dans toutes ses composantes sous-tend fondamentalement notre capacité à modéliser les différents scénarios d'évolution du climat et ses conséquences environnementales, sociales, économiques et politiques. La diversité et les masses de données d'observation et de simulations nécessaires à cette compréhension font du climat et du cycle de l'eau des cas d'écoles de la nécessité d'une infrastructure permettant la convergence de ces données et des moyens d'analyses. Les nombreux enjeux associés à ces questions ont souvent en commun le besoin de mieux caractériser les processus agissant au sein des compartiments ou à leurs interfaces.

<sup>11</sup> L'initiative internationale "4 pour 1000", lancée par la France le 1er décembre 2015 lors de la COP 21 vise un taux de croissance annuel de 0,4% des stocks de carbone du sol par an, dans les premiers 30 à 40 cm de sol, afin de réduire dans l'atmosphère la concentration de CO<sub>2</sub> liée aux activités humaines.  
<https://www.4p1000.org/fr>

<sup>12</sup> <https://www.inrae.fr/actualites/stocker-4-1000-carbone-sols-potentiel-france>

<sup>13</sup> Multilatéralisme et institutions ou agences internationales (Banque Mondiale, FAO, FMI, PNUD, ...)

#### 2.4.1 INTERACTIONS NUAGES - VAPEUR D'EAU - AEROSOLS

Les interactions entre la vapeur d'eau, les aérosols, les nuages et les précipitations demeurent encore aujourd'hui des questions non totalement résolues pour la communauté scientifique et une source d'incertitudes importantes pour la prévision numérique du temps et du climat. La compréhension des processus à partir des observations satellitaires est probablement l'un des enjeux majeurs actuels et les nombreux paramètres devant être décrits de manière simultanée ne peuvent être observés qu'en combinant les informations issues de différents capteurs et différentes plateformes. Cela pose des défis à la fois scientifiques, pour l'interprétation et la simulation cohérente d'observations réalisées sur une vaste plage spectrale, et techniques, compte tenu de la diversité et des volumes de données concernées.

A brève échéance, l'Europe sera dotée de deux nouveaux systèmes d'observation opérationnelle pour la météorologie<sup>14</sup> qui fourniront des observations sans précédent pour l'étude des processus atmosphériques mais qui représenteront un immense défi en termes d'analyses, notamment sur la question des approches multi-capteurs pour lesquelles les agences opérationnelles (EUMETSAT) n'ont pas forcément de filières préexistantes. L'apport évident des données issues de modèles (ERA5 par exemple) doit également être considéré.

**Verrous : Besoin d'un accès centralisé aux données (observations et modèles) et des moyens d'analyse permettant de développer des approches multicapteurs pour une meilleure caractérisation des propriétés de l'atmosphère et d'analyser de très grands volumes de données croisant observations et réanalyses atmosphériques pour l'étude des processus.**

#### 2.4.2 INTERACTIONS CLIMAT ET SURFACES CONTINENTALES

L'interface entre l'atmosphère et les surfaces continentales est un lieu d'échanges importants d'énergie et de matière liquide, solide ou gazeuse. Aussi, comprendre et quantifier les interactions entre surfaces et atmosphère est un enjeu de premier ordre pour anticiper au mieux les conséquences de nouvelles et futures évolutions des surfaces continentales, que ce soit sous l'effet de l'évolution du climat ou de modifications de l'usage des sols. Un des enjeux est aussi de représenter le sol et le sous-sol et les interactions avec la surface de manière beaucoup plus explicite et détaillée que ce qui est fait jusqu'à présent (cf. section sur la zone critique). Par exemple, la profondeur racinaire qui conditionne la reprise de l'eau souterraine par les plantes est encore mal connue, et souvent limitée à un mètre superficiel des sols dans les modèles, ce qui peut amener à sous-estimer l'évapotranspiration réelle, particulièrement en zone tropicale.

**Verrou : Besoin d'un accès centralisé aux données du sol et du sous-sol, de la végétation cultivée et naturelle, des données météo en climat tempéré et tropical.**

#### 2.4.3 ATTRIBUTION DES CHANGEMENTS OBSERVES

Dans les sciences du climat, l'attribution a agrégé une communauté qui a su poser des concepts théoriques et méthodologiques pour mieux y répondre, que ce soit sur l'attribution des événements extrêmes ou sur les tendances climatiques à long terme. Des questions similaires se posent pour l'évaluation des rôles respectifs du changement climatique, d'une part, et de l'usage des sols d'autre part, sur les modifications des flux entrants et sortants d'énergie, d'eau et d'espèces-traces gazeuses,

---

<sup>14</sup> European Polar System – Second Generation » (EPS-SG) et « Meteosat Third Generation » (MTG)

dissoutes ou particulières. La communauté des surfaces continentales est cependant encore peu outillée pour ce sujet du fait de la complexité des systèmes socio-éco-hydro-logiques et de l'interaction des facteurs d'influence.

Aussi, comprendre et quantifier les interactions dans le continuum sol-eau-plante-atmosphère est un enjeu de premier ordre pour anticiper au mieux les conséquences de nouvelles et futures évolutions des surfaces continentales, que ce soit sous l'effet de l'évolution du climat (intensité, intermittence, etc.) ou des modifications liées à l'homme de l'usage des sols, des pratiques culturales, ou du cycle de l'eau (barrages, ouvrages, irrigation, prélèvements d'eau potable). Ces avancées permettront ainsi d'apporter des réponses fondées aux questionnements des décideurs publics.

La complexité particulière de l'analyse des échanges entre surface et atmosphère est principalement liée aux hétérogénéités spatiales de la surface (en nature, en couverture, en type d'usage...), de son évolution temporelle et de la non linéarité des processus.

**Verrous : Besoin d'un accès centralisé aux données (surface/atmosphère) à hautes résolutions spatiales et temporelles.**

#### 2.4.4 DETERMINER LES FLUX AUX INTERFACES

Les processus physiques, chimiques et biologiques qui interviennent aux interfaces entre les différents compartiments du système Terre sont critiques à comprendre pour estimer les flux d'énergie et de matière et permettre d'appréhender et modéliser le système dans son ensemble. Parmi ceux-ci, l'estimation des flux de chaleur et de carbone sont deux défis majeurs, en particulier pour l'anticipation et l'adaptation au changement climatique qui impliquent à la fois des questions scientifiques mais aussi des questions politiques. Où part la moitié du CO<sub>2</sub> émis par l'homme ? Quels sont les paramètres qui contrôlent les échanges naturels de carbone ? Le puits "naturel" (plutôt "forcé") va-t-il se poursuivre au 21<sup>ème</sup> siècle avec le changement climatique annoncé ? Comment évaluer la capacité actuelle et future des écosystèmes terrestres à stocker du carbone ? Quels sont les flux associés à la composante long-terme du cycle du carbone liée à l'altération chimique ?

Est-il possible de vérifier les émissions déclarées par les états, de mesurer ces émissions à plus fine échelle (ville, industrie) pour tracer une activité ? Ces questions sont d'autant plus complexes qu'elles impliquent de caractériser des flux très variables dans le temps et l'espace, en particulier les flux anthropiques d'énergie et de matières. Concernant la capacité de stockage des écosystèmes, il est nécessaire de pouvoir quantifier séparément et à grande échelle les flux de photosynthèse et de respiration.

A l'échelle globale, la télédétection spatiale permet de mesurer des concentrations de CO<sub>2</sub> mais pas directement des flux qui doivent être dérivés des gradients spatiaux-temporels de concentrations en combinant observations et modèles. Pour le CO<sub>2</sub>, comme pour l'ensemble des flux de carbone, force est de constater que les approches diagnostiques et les modèles climat-carbone développés à ces fins produisent des résultats divergents en matière de distribution spatiale, de variations saisonnières et d'intensité des flux biosphériques, autant pour le présent que dans le futur.

L'exemple du carbone illustre parfaitement la complexité de la détermination des flux et la stricte nécessité de pouvoir aborder le problème en disposant simultanément d'observations et de moyens de modélisation à différentes échelles et en intégrant les différents compartiments du système Terre. C'est donc un défi particulièrement fédérateur pour les activités de l'IR Data Terra.

**Verrous : Réussir le passage des observations de grandeurs instantanées (e.g. concentration CO<sub>2</sub>) à leurs gradients spatiaux temporels avec des précisions suffisantes pour permettre la détermination des flux en couplant observations des gradients et modélisation des interfaces.**

#### 2.4.5 BILANS EVAPORATION/PRECIPITATIONS DES OCEANS

L'océan représente 97 % de l'eau totale de la planète et joue donc un rôle clé dans le cycle de l'eau. Outre le fait qu'elle affecte la quantité de vapeur d'eau atmosphérique et donc les précipitations, l'évaporation de la surface de la mer est importante dans le mouvement de la chaleur dans le système climatique. Le bilan hydrique et la salinité sont étroitement liés par l'évaporation et les précipitations. La circulation thermohaline mondiale (Global Conveyer Belt) illustre comment les courants océaniques transportent les eaux de surface chaudes de l'équateur vers les pôles et tempèrent ainsi le climat mondial. La salinité de surface de la mer (SSS) est un traceur clé pour comprendre le cycle de l'eau douce dans l'océan.

L'interface entre l'océan et l'atmosphère est un lieu d'échanges d'eau et de chaleur latente, un enjeu est d'évaluer comment la fonte des calottes glaciaires et le réchauffement des eaux de surface océaniques vont affecter les échanges océan-atmosphère, les bilans évaporation/précipitations au niveau des océans et la circulation thermohaline. Pour cela, il faudra être en mesure de traquer notamment les changements de la salinité de surface de l'océan.

**Verrous : Pouvoir accéder à une combinaison de l'ensemble des données acquises à l'interface océan – atmosphère (in situ, sat) ; développer des algorithmes de traitement des observations et d'assimilation de données, et des produits de visualisation statistiques.**

## 2.5 ALEAS ET IMPACTS SUR LA SOCIETE

Qu'ils soient d'origine naturelle ou induits par les activités humaines, les différents types d'aléas, auxquels est soumis le Système Terre, peuvent avoir un impact considérable sur nos sociétés, en fonction des contextes locaux de développement démographique, urbain et industriel, et dans un contexte plus global de mondialisation. Le suivi et la compréhension de ces aléas nécessitent ainsi une approche intégrée, transverse, croisant données in-situ et de télédétection couvrant l'ensemble des échelles spatiales et temporelles impliquées dans les processus physiques sous-jacents.

### 2.5.1 ALEAS SISMIQUES ET VOLCANIQUES

Avec l'accroissement de la population mondiale et sa densification, de nombreuses zones de failles ou volcans actifs se trouvent aujourd'hui à proximité de grands centres urbains. Les conséquences d'un séisme ou d'une éruption peuvent être aussi bien locales que planétaires (e.g. tsunamis, interruption des circulations aériennes...). Une meilleure quantification des aléas sismiques et volcaniques demeure donc un enjeu majeur, en amont de l'évaluation des risques associés et de leur gestion. Un défi clé est de fournir un suivi spatio-temporel à haute résolution et précision des déformations associés à ces objets naturels. Ce suivi doit pouvoir se faire à l'échelle des roches, des « objets » failles et volcans, comme à l'échelle des continents, la physique des événements telluriques impliquant les mouvements internes de la Terre comme des circulations de fluide à toute petite échelle. Les progrès en instrumentation et télédétection de ces dernières décennies (progrès techniques ou méthodologiques) ont permis des avancées et découvertes majeures : découverte des séismes lents sur les failles remettant en cause notre vision du cycle sismique, prédiction de certaines

éruptions et suivi des émissions par exemple. La profusion des données à disposition fait cependant apparaître de nouveaux verrous, pour l'analyse jointe de données de disciplines variées, l'amélioration des rapports signal sur bruit (propres à chaque discipline mais interconnectés, le bruit en géophysique interne pouvant être le signal en physique de l'atmosphère et inversement), ou encore la modélisation mécanique intégrant toutes les échelles.

L'enjeu pour l'aléa sismique est de pouvoir fournir des modèles probabilistes globaux, évolutifs dans le temps, en s'appuyant sur la combinaison des données tectoniques, sismologiques et géodésiques (déformation et gravimétrie). Pour les aléas volcaniques, la surveillance et les systèmes d'alerte nécessitent une disponibilité en temps réel ou proche (NRT) des observations spatiales à la fois pour l'imagerie haute résolution mais aussi pour les produits (gaz, SO<sub>2</sub>, aérosols, ...). Le développement des capacités de traitement et de diffusion NRT des données de télédétection pour les zones surveillées doit se faire en lien avec la mise à disposition des autres données d'observation au sol et in-situ, tant pour la composition atmosphérique que pour les mesures de température, déformation et de sismicité. La création d'une plateforme multidisciplinaire d'accès et de visualisation de données pour la surveillance et l'étude des volcans regroupant capacités d'observation et de modélisation serait un apport significatif par rapport à des systèmes existants<sup>15</sup>.

**Verrous : plateforme multidisciplinaire d'accès et de visualisation de données ; analyses conjointes, amélioration des rapports signal/bruit, séparation des signaux, modélisation, calculs en temps réel (NRT), diffusion de l'information NRT**

## 2.5.2 INTENSIFICATION DES CYCLES HYDROMETEOROLOGIQUES

Le changement climatique provoque une intensification du cycle hydrométéorologique qui induit des sécheresses comme des inondations. Les besoins opérationnels sont réels mais les aléas hydrométéorologiques sont encore mal identifiés dans les séries existantes et mal prévus dans les services opérationnels. Il est nécessaire aujourd'hui de mieux connaître leur variabilité à l'échelle pluriannuelle et de renseigner leur évolution, dans le contexte du changement climatique. Identifier les lois des extrêmes dans les séries chronologiques pose un défi théorique dans un climat changeant. Il est pourtant nécessaire de connaître ces lois pour évaluer les modèles, mais aussi pour prévoir les dimensionnements des barrages et des ouvrages et prévenir les inondations. Ces dernières années, des avancées significatives dans la prévision des événements de pluie intense ont été permises par la mise en service opérationnel de modèles de prévision non hydrostatiques à quelques kilomètres de résolution (comme le modèle AROME de Météo-France) permettant la prévision des débits de bassins versants à réponse rapides (comme en Méditerranée). Cependant pour les petits bassins, une erreur de quelques dizaines de kilomètres peut conduire à prévoir la crue sur le bassin voisin. L'enjeu est maintenant d'intégrer les impacts dans les modèles couplés hydrométéorologiques. Il faut donc pouvoir accéder à des informations qui relèvent des enjeux matériels et humains.

**Verrous : pour mieux prévoir et modéliser les pluies intenses et crues rapides, besoin de mieux quantifier et réduire les incertitudes des prévisions et des projections régionales climatiques en améliorant notre compréhension des processus en jeu et de leur représentation dans les modèles numériques, en développant l'assimilation de données dans les domaines pauvres en observations assimilées, et en développant des méthodes de**

<sup>15</sup> Par exemple: [www.mounts-project.com/](http://www.mounts-project.com/) ou [//www.mirovaweb.it/](http://www.mirovaweb.it/)

---

**prévision d'ensemble permettant de quantifier la prévisibilité d'un événement. Besoin de disposer de données d'observations à des échelles variées et sur le long terme.**

### 2.5.3 RESSOURCES MINÉRALES ET ÉNERGETIQUES

La transition vers des énergies décarbonées est inéluctable mais requiert des besoins nouveaux et considérables en ressources minérales (métaux, terres rares, granulats...). Réussir cette transition nécessite donc de relever plusieurs défis. Quelles que soient les technologies employées pour la production et l'exploitation des ressources, un meilleur suivi spatio-temporel de leur impact environnemental est indispensable : analyses in-situ ou par télédétection des pollutions locales, suivi par géodésie (sol ou spatiale) et sismologie de la déformation et de la stabilité des sols dans les zones d'exploitation, optimisation des réseaux d'éoliennes, etc. Des approches innovantes sont également à développer pour la recherche de nouvelles ressources et leur exploitation durable minimisant l'impact sur les sociétés et le milieu. Un exemple est le développement de nouveaux systèmes d'imagerie (hyperspectrale...) et de traitement d'images pour l'exploration depuis l'espace (recherche de lithium...). Un autre exemple concerne les développements associés à la mine urbaine.

**Verrous : Mieux quantifier et réduire ces aléas requiert donc la mobilisation de plusieurs compartiments du système Terre (voir aussi section 2.1.5), un des verrous actuels étant de pouvoir croiser, entre autres, des données géologiques et environnementales in-situ, des données d'imagerie par télédétection, voire pour certains types de ressources, nos connaissances des milieux (urbain, marin...).**

## 3 BESOINS SUR LES DONNÉES

### 3.1 IDENTIFICATION DES VERROUS COMMUNS

---

L'illustration de défis transverses aux interfaces des pôles thématiques a permis d'identifier des verrous qui restent liés à l'accessibilité et l'utilisation des données. L'examen de ces verrous montre que la plupart sont communs à l'ensemble des communautés scientifiques, quel que soit le compartiment du système Terre étudié. Les deux enjeux majeurs concernent d'une part la mise à disposition des données et d'autre part la capacité de les exploiter.

La première priorité est de pouvoir accéder à l'ensemble des données du système Terre indépendamment de leurs sources : données satellites, observation in situ, mesures et expérimentations au laboratoire, banques de données, sorties de modèles. Ces données doivent être disponibles et accessibles en qualité et résolution suffisantes. La seconde priorité est de pouvoir combiner ces données à différentes échelles de temps et d'espace pour favoriser l'approche intégrée. Il y a un besoin critique de développer une offre d'outils pour l'acquisition, la mise à disposition et l'exploitation des futurs lacs de données d'origine et de nature diverses. Pour tirer le meilleur parti du flux massif et croissant de données, il est maintenant nécessaire de développer des centres de données et de services d'accès, de traitement et de visualisation de données. Il sera ainsi possible de mettre en œuvre une approche intégrative, prenant en compte l'ensemble des processus des interactions surfaces continentales / terre solide / atmosphère / océan à toutes les échelles.

Ces besoins constituent un challenge pour l'IR Data Terra qui a ainsi identifié les besoins stratégiques prioritaires sur les données et les défis méthodologiques qui en découlent.

### 3.2 LES PRINCIPES FAIR

L'IR Data Terra souscrit complètement aux principes de la Science Ouverte<sup>16</sup>, qui visent à rendre les données scientifiques FAIR<sup>17</sup>, à savoir qu'elles soient toujours facilement découvrables, accessibles, interopérables et réutilisables. La FAIRisation des données est donc un enjeu prioritaire de DataTerra. Parmi la diversité des dépôts de données accessibles via le portail de Data Terra et des pôles de données thématiques, les services les plus matures, souvent associés aux services d'observations labellisés, ont déjà atteint un niveau élevé de FAIRisation. Mais il demeure encore de nombreux jeux de données, qui ne respectent pas ce principe.

Pour l'étude du Système Terre, par le biais d'observations permanentes et temporaires, le principe de base est que les données doivent être découvrables et accessibles, afin d'en permettre l'interopérabilité et la réutilisation par toutes les communautés scientifiques. Ces notions se décrivent selon :

- **Découvrable** : Être découvrable signifie que les données sont décrites par des métadonnées riches, respectant des normes internationales reconnues, dans les portails de découverte de données et ont un identifiant persistant unique (ex. DOI).
- **Accessible** : Être accessible signifie que les données peuvent être facilement obtenues par une intervention humaine ou une interaction de machine à machine selon des protocoles définis. La nature des données et les conditions d'accès et de réutilisation doivent être clairement définies (provenance, format, licence, ...). Les métadonnées, au minimum, doivent être accessibles afin que le jeu de données reste découvrable et permettre de contacter l'auteur du jeu de données pour une demande d'accès aux données. Les informations relatives à la licence doivent pouvoir être mises à jour par l'utilisateur. Data Terra privilégie les licences qui imposent l'attribution de la provenance des données.

*À ces exigences sont associés des principes standard pour la gestion des dépôts de données et des services, y compris la sécurité des données et la continuité et la robustesse des services. C'est pourquoi Data Terra encourage la certification des dépôts de données.*

- **Interopérable** : L'interopérabilité est obligatoire pour l'intégration de différents jeux de données sur un portail unique. Cela implique l'utilisation d'un vocabulaire aux normes et d'un langage de programmation commun. Le GT Technique de Data Terra vise à harmoniser ces pratiques à l'ensemble des pôles.
- **Réutilisable** : L'application des trois principes précédents permet aux données d'être réutilisées de façon optimale. Il est impératif que les données, les métadonnées et les licences d'utilisation soient extrêmement détaillées. Le producteur doit être averti du téléchargement de ses données.

L'application des principes FAIR aux jeux de données portés par Data Terra encouragera les travaux inter-domaines, pour lesquels les interprétations multi-domaines impliquent que des chercheurs non spécialistes soient capables d'utiliser, au moins dans une certaine mesure, les données d'un autre domaine.

<sup>16</sup> [www.science-ouverte.cnrs.fr/wp-content/uploads/2019/11/Plaquette\\_Science-Ouverte\\_18112019.pdf](http://www.science-ouverte.cnrs.fr/wp-content/uploads/2019/11/Plaquette_Science-Ouverte_18112019.pdf)

<sup>17</sup> FAIR : Findability, Accessibility, Interoperability, Reuse - [www.go-fair.org/fair-principles](http://www.go-fair.org/fair-principles)

### 3.3 TRAITEMENTS, CALCULS ET SERVICES

La diversité et le volume des données acquises via l'observation ou générées par la modélisation posent des problèmes fondamentaux aussi bien d'un point de vue méthodologique que technique. L'interprétation, la mise en relation et l'intégration dans des modèles de données de natures très différentes (ex : croisement d'informations géophysiques avec des mesures biochimiques) posent par exemple des problèmes méthodologiques, même lorsque les volumes de données sont limités.

De surcroît, la croissance combinée des données issues de l'observation spatiale, de la modélisation et la simulation numérique amène progressivement à devoir considérer et manipuler des volumes de l'ordre de l'exascale. Les évolutions de ces sources de données conduisent à une convergence des infrastructures soutenant la science et mêlant données d'observation et simulations.

Pour fixer les ordres de grandeur, les données de réanalyse du Centre Européen pour les Prévisions Météorologiques à Moyen Terme (ECMWF<sup>18</sup>, ERA-5), représentent à elles seules 9 pétaoctets (Po) de données. L'archive des observations des différents satellites Sentinel, quant à elle, a atteint en 2018 près de 10 Po de données, dont 5 Po produites pour la seule année 2018. L'arrivée prochaine des données de Meteosat Troisième Génération (MTG), avec en particulier celle du sondeur infrarouge, ne fera qu'amplifier cette tendance.

Les verrous techniques pour l'accessibilité, la manipulation et le traitement de tels volumes de données sont indissociables des questions de coûts d'infrastructures. La complexité et l'hétérogénéité des systèmes informatiques augmentent alors même qu'il devient de plus en plus coûteux de déplacer les données. Dans le même temps, la nécessité d'une expertise scientifique et de communautés utilisatrices localisées proches des données reste importante pour mobiliser les différents acteurs (utilisateurs et financeurs). Cet équilibre peut être atteint via une architecture distribuée reposant sur des centres disposant de la taille critique nécessaire pour répondre au besoin de mutualisation tout en garantissant une dynamique locale et une émulation au niveau national. Cette logique a conduit à l'apparition de nouvelles plateformes distribuées de services supportées par un continuum d'infrastructures pour accélérer la logistique de ces données et faciliter l'accès à ces ressources. C'est la stratégie retenue pour la construction de l'IR Data Terra.

#### 3.3.1 ANALYSE DE DONNEES

L'analyse des très larges volumes de données ou de données diverses hébergées dans différents centres nécessite le développement et la réalisation d'un ensemble de services et d'outils interdisciplinaires basés sur le calcul haute performance (HPC). C'est un enjeu majeur soutenu par la communauté européenne, au travers notamment le projet H2020 PHIDIAS, auquel contribue la communauté de l'IR Data Terra, dont l'objectif est de faire émerger des premières solutions concrètes pour ce type de services de traitements massifs. Les services de découverte et de logistique des données sont naturellement à concevoir dans l'esprit FAIR. Les moyens HPC et HPDA (High Performance Data Analysis) doivent être mis en place quant à eux pour répondre au besoin croissant d'analyses et de traitements à la demande. L'essor des services de traitements à la demande et de mise en place d'environnements de recherche virtuels<sup>19</sup> repose par ailleurs sur :

- le libre accès à des services de calcul haute performance (HPC) standardisés ;
- le développement de nouveaux modèles de traitement de données pouvant être couplés aux moyens HPC ;

<sup>18</sup> ECMWF is the European Centre for Medium-Range Weather Forecasts - [www.ecmwf.int](http://www.ecmwf.int)

<sup>19</sup> Virtual Research Environment (VRE)

- le déploiement de méthodes de traitement disponibles sous forme de services.

L'indispensable reproductibilité des résultats d'analyse impose enfin que les chaînes de traitement et d'analyse puissent être préservées, en particulier si les volumes de données générés sont conséquents et que le rapport coût de stockage/coût de traitement devient important. Il devient alors plus économique de sécuriser les chaînes de traitement et de les rejouer à la demande plutôt que de stocker des résultats volumineux.

### 3.3.2 METHODES MATHEMATIQUES ET INTELLIGENCE ARTIFICIELLE

Sur un plan technique, en dehors des aspects liés à la fourniture d'accès et à la possibilité d'opérer des traitements massifs sur des données de plus en plus variées et volumineuses, Data Terra doit aussi permettre d'ouvrir les bases de données à de nouvelles communautés pour lesquelles les mégadonnées (big data) sont un objet d'étude en soi. Il faut favoriser les interactions entre les communautés issues de la « géophysique » et celles de l'informatique, et plus particulièrement celle de l'Intelligence Artificielle (IA). Les techniques sophistiquées notamment issues de l'IA sont particulièrement performantes pour le traitement des séries temporelles, comme par exemple la séparation des signaux provenant de différentes sources, la détection d'anomalies ou de précurseurs ou enfin l'amélioration du rapport signal/bruit.

### 3.3.3 FOUILLE DE DONNEES

L'augmentation de la quantité et des volumes des données conduit à des difficultés d'appropriation des données par les équipes scientifiques, non seulement au niveau de leur découverte, mais aussi dans leur réel contenu, leur variabilité ou leurs interdépendances. Cela induit la nécessité d'une offre de service de visualisation en ligne (par exemple cartographie, séries temporelles) d'analyses géostatistiques, de modélisation inverse.

Comme il devient de plus en plus difficile de télécharger les données sur son poste de travail, il s'agit ici de proposer des *laboratoires virtuels* d'analyse et de fouille de données, mis en œuvre là où sont stockées les données, sur la base de logiciels existants ou de bibliothèques (par exemple PANGEO, ...). Ceci inclut le prototypage de nouveaux algorithmes (par exemple, par le développement de script Python, R, ...).

Ces services nécessitent des moyens informatiques conséquents et des environnements de recherche virtuels (VRE) : lac de données, espace de stockage de travail (résultats intermédiaires), moyens de calcul (CPU, GPU, ...), nœuds pour les liens interactifs avec les utilisateurs. Le concept de lac de données regroupe une famille d'outils de stockage pour fouiller des ensembles très volumineux de données provenant de sources multiples et hétérogènes.

Les produits en aval peuvent prendre de nombreuses formes, par exemple :

- Données prétraitées : les données brutes ayant été nettoyées de ce qui peut être considéré comme du bruit ou des données fausses (par exemple les séries temporelles de GPS corrigées des effets atmosphériques ; ou la qualification des données issues de capteurs in situ par rapport à des normes) ;
- Produits secondaires dérivés des données brutes (par exemple les localisations des tremblements de terre, flux d'évapotranspiration, production d'indicateur de qualité des eaux) ;
- Produits impliquant un mélange de modélisation et de données brutes (par exemple les modèles de hauteur de vagues).

---

## 4 DEFIS METHODOLOGIQUES

---

Les défis méthodologiques actuels sont liés à l'explosion du nombre et du flux de données sol et spatiales et à **l'imbrication et la multiplicité des échelles spatiales et temporelles** impliquées dans l'étude du Système Terre. Ces défis sont communs à différentes communautés et requièrent des passerelles et des interactions fortes entre elles, que Data Terra doit contribuer à construire. L'accessibilité simplifiée aux données et à des moyens de calcul communs adaptés au traitement massif de données volumineuses est une première base (sections 3.1 et 3.2). D'autres défis concernent plus spécifiquement les étapes de traitement et d'analyse des données (amélioration de la précision nécessitant l'apport de disciplines complémentaires), les méthodes de combinaison et d'inversion conjointes de données d'origine, de résolution et de sources d'erreur variées, et les approches de modélisation physique qui doivent pouvoir intégrer les différentes échelles.

### 4.1 INCERTITUDES ET SEPARATION DES SIGNAUX

Pour extraire les observables ou les paramètres utiles à un domaine de recherche, l'identification des **sources d'erreur**, leur correction éventuelle et la quantification des **incertitudes résiduelles** sont essentielles. Indépendamment de l'amélioration de la compréhension des phénomènes physiques à l'origine des erreurs, il apparaît important de développer de **nouvelles méthodes de séparation de signaux** dans les données, reposant sur le croisement des connaissances propres à chaque domaine de recherche, ou sur l'utilisation de **méthodes de fouille de données et d'intelligence artificielle** capables de reconnaître la signature spatio-temporelle spécifique à chaque signal. En Terre Solide par exemple, l'obtention de séries temporelles de déplacement précises, issues de données spatiales, nécessite de mieux séparer les signaux de déformation du sol recherchés (potentiellement d'origines diverses : tectonique, géodynamique ou anthropique...) du bruit, associé aux variations de la charge hydrologique ou aux délais de propagation des ondes dans l'atmosphère. Pour repousser encore les limites de détection de certaines observables, mêmes localement, il est également important d'améliorer certains modèles globaux (par exemple les systèmes de référence en géodésie ou les modèles atmosphériques globaux), et de les rendre disponibles pour les non-spécialistes.

### 4.2 ASSIMILATION ET COMBINAISON DE DONNEES MULTI-TEMPORELLES ET MULTI-DISCIPLINAIRES

L'analyse de certains phénomènes à différentes échelles spatiales et temporelles, la comparaison avec des modèles existants ou l'élaboration de nouveaux modèles, impliquent fortement de pouvoir utiliser conjointement les données spatiales et non spatiales (sol – in situ). Ceci nécessite un accès aisé aux données par les utilisateurs et développeurs d'applications scientifiques et des moyens de traitement facilitant les analyses conjointes. En termes d'analyse, les étapes de colocalisation de l'ensemble des données, d'interpolation, d'inversion, de modélisation, devront encore être améliorées. Ceci permettra de mieux prendre en compte la spécificité de chaque type de données en termes de résolution spatiale et/ou temporelle et d'incertitudes, et de renforcer ainsi leur complémentarité.

### 4.3 SCIENCES PARTICIPATIVES

Les années récentes ont vu l'émergence de nouveaux types de projets pluridisciplinaires, au croisement de la recherche scientifique, de la démocratie participative et de l'innovation

technologique. Ces projets proposent des modalités d'actions concrètes en faveur d'une meilleure surveillance de l'environnement et d'une transition écologique inclusive. Mais le cycle de vie d'un projet de science participative soulève des problématiques inédites, des défis singuliers.

Les sciences participatives doivent permettre à l'avenir d'acquérir des données en masse utilisables par les chercheurs. Elles doivent également favoriser la communication vers le citoyen et un retour vers l'expert. Mais la réussite de ce type de projets nécessite la maîtrise et la mise en œuvre de processus dont beaucoup doivent être développés, partagés, et mutualisés : le contrôle qualité, dont l'étalonnage (lorsque des capteurs sont utilisés) et la correction de la dérive dans le temps, la production et la distribution de données validées, l'animation du projet, la restitution des résultats ... Le principe FAIR prend en général une grande importance dans ces projets. Les pôles de données peuvent être la référence pour beaucoup des processus mis en œuvre dans ces projets.

## **5 BESOINS EN FORMATION ET COMMUNICATION**

Le projet vise à promouvoir l'utilisation et l'analyse de données issus des observations et de la modélisation dont le volume et la complexité sont croissants. La gestion de telles bases de données est un domaine dynamique, où les outils informatiques sont en constante évolution et encore peu familiers à la plupart des scientifiques, et quasi inaccessibles à la société civile. Cela entraîne de nouveaux défis de formations pour les communautés concernées.

### **5.1 FORMATION INTERNE DES ACTEURS DE DATA TERRA :**

Le projet permettra l'utilisation et l'analyse de données d'origines multiples. Cela impose d'être rigoureux dans la production d'ensembles de données de plus en plus volumineux et complexes. L'enjeu des centres de données et de service (CDS) de l'IR Data Terra sera de maintenir à long terme l'assurance de la qualité et de l'accessibilité des données archivées par un maintien à niveau des protocoles d'archivage, de conservation et de moissonnage des données pour accompagner les évolutions de format, de support de stockage, de protocoles. Ils doivent aussi s'inscrire dans le respect de la démarche FAIR et des certifications reconnues (ex. Core Trust Seal). Cela implique une formation continue des personnels des centres de données et de service. Cette formation se réalisera par des ateliers techniques internes à l'IR Data Terra, qui permettront des échanges d'expérience et des travaux pratiques. Les CDSs profiteront aussi des formations dispensées par d'autres instances nationales et internationales telles RDA-France, EOSC (séminaires, webinaires).

### **5.2 FORMATION DE LA COMMUNAUTE SCIENTIFIQUE :**

La structuration de l'IR Data Terra donnera accès à un flux de données d'observation et de modélisation sans précédent et à des outils virtuels innovants. Or, de nombreux scientifiques n'ont pas encore l'habitude de tels usages, voire n'ont pas de formation approfondie en science des données et en programmation. L'enjeu de formation de la communauté scientifique porte sur trois volets :

#### **5.2.1 DE LA RECOLTE A LA FAIRISATION DE LA DONNEE**

Améliorer la formation des scientifiques à la conception et à la mise en œuvre de plans de gestion des données (DMP) efficaces dans le respect des principes FAIR ;

Ces formations pourront être réalisées en auto-formation grâce à des documents, des vidéos et des webinaires mis à disposition sur les sites de l'IR Data Terra et des pôles de données.

### 5.2.2 APPRENTISSAGE DE L'UTILISATION D'ENSEMBLES DE DONNEES D'OBSERVATION

Le but est de développer une formation spécialisée pour fournir une approche pratique du système pour informer sur les données et les produits disponibles, et les modes de découverte des données comme service de visualisation en ligne, cartographie, séries temporelles. Il sera aussi nécessaire de former sur les conditions d'utilisation des données et le respect des sources (DOI, licences creative commun).

### 5.2.3 APPRENTISSAGE DE L'UTILISATION DES SERVICES (VRE) A LA DEMANDE

Il y aura deux niveaux à considérer :

- Initiation : renforcer les capacités des scientifiques à adopter dans leur travail l'utilisation des outils mis à disposition (travail à distance, cloud, analyses géostatistiques, combinaison des données de sources différentes pour une même zone, ...)
- Expert : acquérir de l'expérience sur la façon d'analyser et d'interpréter les données avec des options données sur les langages de codage (Python).

Ces formations pourront être organisées par l'IR Data Terra ou en collaboration avec les Universités impliquées dans l'IR.

## 5.3 COMMUNICATION AUTOUR DE NOTRE OBJET D'ETUDE :

L'IR Data Terra et les pôles thématiques poursuivront :

- leur démarche volontaire d'informations de la communauté scientifique par des participations à des colloques nationaux et par des séminaires dans les unités de recherche. Ces participations ont pour but d'informer sur l'existence de l'IR Data Terra, de son offre de service.
- la diffusion d'actualités sur leurs pages web et via les réseaux sociaux (twitter, ...).

Une attention particulière devra être portée vers le transfert de connaissance vers la société civile. De nombreuses données ont un intérêt pour les gestionnaires et les politiques, et même le citoyen, qui n'ont le plus souvent pas l'habitude ni les bases scientifiques pour naviguer sur le portail de données et de services. Un effort de communication devra être mis en place par l'IR Data Terra pour présenter des données de façon simple et synthétique. Par exemple, la climatologie des eaux côtières peut être déclinée dans une version publique avec une présentation des données de l'année et comparaison avec les données historiques ou la comparaison de différents sites. Cette action de communication sur les données est très importante dans un contexte de changements globaux pour informer de façon factuelle sur l'évolution de la Terre et de l'Environnement.