
DMP du projet "TRACCS PC2 : Brokerage of data and methods"

Plan de gestion de données créé à l'aide de DMP OPIDoR, basé sur le modèle "Science Europe : modèle structuré" fourni par Science Europe.

Renseignements sur le plan

Titre du plan	DMP du projet "TRACCS PC2 : Brokerage of data and methods"
Livrable	
Version	Version initiale
Domaines de recherche (selon classification de l'OCDE)	Earth and related environmental sciences
Langue	fra
Date de création	2024-03-22
Date de dernière modification	2024-07-03
Identifiant	
Licence	Nom Creative Commons Attribution 4.0 International
	URL http://spdx.org/licenses/CC-BY-4.0.json
Plans de gestion en lien avec le projet	<ul style="list-style-type: none">Plan de Gestion de Donnée d'ESPRI : https://cloud.ipsl.fr/index.php/s/CgkHcy6nfBongEM

Renseignements sur le projet

Titre du projet	TRACCS PC2 : Brokerage of data and methods
Acronyme	INVEST
Résumé	<p>La quantité de données générées par les modèles climatiques devrait augmenter énormément dans les dix prochaines années avec la puissance de calcul disponible, l'avènement de nouvelles architectures et hausses de la puissance de calcul permettant une résolution plus fine et un plus grand nombre de simulations. Ces grands volumes de données à partager nécessitent des changements de paradigme (stockage objet, compute-to-data, traitement massivement parallèle, etc.). Le projet ciblé #2 INVEST du programme de recherche TRACCS vise à fournir une infrastructure partagée robuste et durable pour distribuer les données et les logiciels climatiques aux acteurs scientifiques et sociétaux selon quatre axes principaux :</p>

1. Préparation pour mise en ligne de jeux de données descendus d'échelle et d'indicateurs sectoriels tirant parti des travaux menés dans le programme TRACCS notamment la connaissance des besoins des utilisateurs (PC1 et PC3), les méthodes de descente d'échelle physique et statistique (PC 4 et 10), le traitement des extrêmes (PC4).
2. Mise à jour des infrastructures nationales DRIAS et ESPRI de distribution et d'accès aux données avec les nouvelles projections régionales et mondiales. Le portail DRIAS et son extension "DRIAS Eau" ouverte fin 2022 évolueront pour pouvoir proposer un portail contextuel pouvant servir un plus grand nombre de besoins sectoriels (agriculture, énergie, tourisme ...). Cela implique la création et la distribution de nouveaux indicateurs, l'accès à des bases de données distribuées, une mise à jour des outils graphiques, ainsi que la mise à disposition d'un atlas interactif fournissant des informations à haute résolution en complément des graphiques standard pouvant être utilisés directement dans les rapports et les évaluations.
3. Fourniture des services haut de gamme pour les utilisateurs avancés qui s'appuient sur les développements réalisés dans le programme TRACCS et GAIA-DATA. Des couches spécifiques seront également interfacées avec des infrastructures européennes plus génériques telles que Climate4Impact ou EOSC, EGI et EUDAT. En particulier, nous ajouterons une couche de préparation des données et de technologie spécifique aux sorties des modèles climatiques pour un traitement efficace, basé sur le cloud, des données de masse et des analyses de données prêtes à l'emploi (par exemple, en utilisant les technologies Zarr et Dask).
4. Définition et fourniture d'une infrastructure partagée conforme à la norme FAIR afin d'offrir un meilleur accès aux données, aux méthodes statistiques innovantes et aux codes associés qui ont été développés dans le cadre de TRACCS. Nous contribuerons à la définition de normes pertinentes avec la Research Data Alliance (RDA). La synergie avec Copernicus/DestinE sera essentielle pour assurer une bonne intégration dans les initiatives européennes.

Sources de financement

- Agence Nationale de la Recherche : ANR-22-EXTR-0003

Date de début

2023-05-01

Date de fin

2031-04-30

Partenaires

- Météo-France <https://ror.org/0233st365>
- Sorbonne Université <https://ror.org/02en5vm52>
- French National Centre for Scientific Research <https://ror.org/02feahw73>
- CNRS délégation Occitanie Ouest

Produits de recherche :

1. Simulations et indicateurs climatiques (Jeu de données)
2. Outils et services à la donnée (Logiciel)

Contributeurs

Nom	Affiliation	Rôles
Ben Nasser Atef - https://orcid.org/0000-0001-6948-8735		<ul style="list-style-type: none"> • Collecteur de données (Datasets INVEST)
KERDONCUFF Maryvonne	Météo France - https://ror.org/0233st365	<ul style="list-style-type: none"> • Collecteur de données (Datasets INVEST, Datasets INVEST) • Gestionnaire de données (Datasets INVEST, Datasets INVEST, Datasets INVEST, Datasets INVEST) • Personne contact (Datasets INVEST)
LEVAVASSEUR Guillaume - https://orcid.org/0000-0002-0801-0890	Sorbonne Université - https://ror.org/02en5vm52	<ul style="list-style-type: none"> • Curateur de données (Datasets INVEST, Datasets INVEST) • Gestionnaire de données (Datasets INVEST, Datasets INVEST) • Personne contact (Datasets INVEST) • Responsable de l'entité (Datasets INVEST, Datasets INVEST) • Responsable du plan
PAGE Christian - https://orcid.org/0000-0001-7743-9570	French National Centre for Scientific Research - https://ror.org/02feahw73	<ul style="list-style-type: none"> • Membre du projet (Outils INVEST, Outils INVEST, Outils INVEST) • Personne contact (Outils INVEST) • Responsable de l'entité (Datasets INVEST)
SOUBEYROUX Jean-Michel	Météo-France - https://ror.org/0233st365	<ul style="list-style-type: none"> • Coordinateur du projet • Personne contact (Datasets INVEST, Outils INVEST)
troussellier laurent - https://orcid.org/0009-0000-9361-0980		<ul style="list-style-type: none"> • Personne contact (Datasets INVEST)

Droits d'auteur :

Le(s) créateur(s) de ce plan accepte(nt) que tout ou partie de texte de ce plan soit réutilisé et personnalisé si nécessaire pour un autre plan. Vous n'avez pas besoin de citer le(s) créateur(s) en tant que source. L'utilisation de toute partie de texte de ce plan n'implique pas que le(s) créateur(s) soutien(nen)t ou aient une quelconque relation avec votre projet ou votre soumission.

DMP du projet "TRACCS PC2 : Brokerage of data and methods"

Description des données et collecte ou réutilisation de données existantes

Simulations et indicateurs climatiques

[Description générale du produit de recherche](#)

Nom	Simulations et indicateurs climatiques
Description	<p>Le projet ciblé (PC) 2, ou INVEST, du programme de recherche TRACCS, vise à sélectionner, préparer et distribuer vers un large public, les données et produits climatiques élaborés dans les autres PC de TRACCS, ou provenant d'autres archives de référence utiles pour les services climatiques, telles que :</p> <ul style="list-style-type: none"> • Des simulations climatiques globales et régionales des exercices de référence CMIP et CORDEX ; • Des simulations climatiques de référence corrigés de leurs biais ; • Des indicateurs pré-calculés à partir des simulations climatiques. <p>Ces données sont exclusivement des données numériques tabulaire. Le format d'origine est HDF5 (Hierarchical Data Format), les données selon le plateau technique de destination pourront être soit :</p> <ul style="list-style-type: none"> • Conservées dans leur format d'origine et notamment sa déclinaison NetCDF (https://www.unidata.ucar.edu/software/netcdf/) ; • Converties dans un format objet "cloud-read" type Zarr (https://zarr.dev/) • Ou se présenter sous la forme d'extractions (géographiques ou temporelles) ou statistiques (séries temporelles, moyennes, etc.) au format CSV. <p>La traçabilité des données sera assurée par l'utilisation :</p> <ul style="list-style-type: none"> • D'identifiants persistants (PID) automatiquement attribués à chaque fichier ou ensemble de fichiers rassemblés en "jeux de données" ou "datasets" ; • D'identifiants d'objets numériques (DOI) qui pourront être déposés sur des ensemble de données particuliers à la demande des utilisateurs auprès d'entrepôts de données certifiés CoreTrustSeal comme ESPRI (Centre de Calcul et de Donnée de l'IPSL) ou le portail EasyData ; • Des services d'errata (comme le service Errata de ESGF) qui assureront la documentation des différentes versions de jeux de données et leur généalogie. <p>Ces objets numériques seront documentés directement au sein des métadonnées (attributs netCDF ou Zarr).</p> <p>Enfin, les données collectées dans le cadre du projet INVEST sont uniquement des données environnementales et ne sont soumises à aucune exigence de confidentialité ou d'éthique. Néanmoins, les collecteurs de données du projet s'engagent à suivre un code de conduite disciplinaire afin de fournir toutes les informations qui se réfèrent directement à l'origine des données, aux règles de partage et à la citation des données.</p>
Type	Jeu de données
Workpackage	WP1
Mots clés (texte libre)	
Langue	anglais
Date de publication	
Contient des données personnelles ?	Non
Contient des données sensibles ?	Non
Prend en compte des aspects éthiques ?	Non

Est-ce que des données existantes seront réutilisées ?

Justification

Le projet INVEST mettra à disposition des données issues de simulations climatiques existantes dont les données sont ouvertes et libre d'accès. La réutilisation de ces données est tout à fait pertinente dans le contexte de TRACCS et même nécessaire afin de valider les avancées des autres projets ciblés. La redistribution de ces données existantes n'engage aucune contrainte car sous licence libre (CC-BY 4.0).

Notamment, le WCRP CMIP Infrastructure Panel (WIP) a approuvé un assouplissement de la licence de CC-BY-SA à CC-BY en 2022 pour le données CMIP5, CMIP6 et CORDEX qui seront réutilisés dans le projet INVEST.

Données réutilisées

- Archive CMIP5 : <https://pcmdi.llnl.gov/mips/cmip5/>
- Archive CMIP6 : <https://pcmdi.llnl.gov/CMIP6/>
- Archive CORDEX : <https://cordex.org/>

Comment seront produites/collectées les nouvelles données ?

Nom de la méthode

Description

Les travaux du projet INVEST s'appuient sur le concours de deux comités transverses, dont un Comité Scientifique (CS) qui assure l'appui scientifique en lien avec les leaders des différents projets ciblés de TRACCS, complétés le cas échéant par des experts représentant d'autres producteurs de données sur les impacts climatiques. Une des activités du CS sera donc notamment d'encadrer la préparation et la collection des jeux de données et des indicateurs au sein du projet INVEST.

Chaque PC TRACCS propriétaire/fournisseur de données sera chargé de respecter les instructions du PGD. Les métadonnées, nom des fichiers et leur arborescence respecteront les normes et bonnes pratiques imposées pour chaque type de donnée et décrits dans leurs "Data Reference Syntax" respectives.

Le gestionnaire de données du PC2 supervisera la mise en œuvre du plan de gestion des données en mettant à jour le PGD, en fournissant les meilleures pratiques ou des fiches d'information pour faciliter la gestion des données par les partenaires.

Un contrôle qualité des métadonnées sera défini de façon commune et appliqué par chaque plateau technique (ou "plateforme") contribuant au projet INVEST afin de s'assurer que les données à collecter sont bien conforme au standards décrit en section 2.

Le processus de collecte reposera essentiellement sur des transferts (migration) ou copies (réplication) manuelles avec des mises à jour régulières ou "à la demande" des autres projets ciblés, en utilisant les logiciels et protocoles en vigueur sur chaque plateforme.

Une stratégie commune de curation des archives définira les critères pouvant amener à remplacer ou supprimer les versions obsolètes des ensembles de données au sein de chaque plateforme. La provenance des données sera inscrite au sein de chaque fichier via un ensemble d'attributs établis par le standard de métadonnées choisi afin de définir l'origine de la donnée (projet, modèle, institut, post-processing, etc.).

Nature des données

Simulation

Equipements, plateaux techniques utilisés

- ESPRI - Centre de Calcul et de Données de l'IPSL : <https://espri.ipsl.fr/>
- DRIAS - Portail de diffusion de Météo France pour les projections climatiques ajustées sur les territoires nationaux : <https://www.drias-climat.fr/>
- Climate4Impact - Portail de diffusion des projections climatiques, connecté à l'Earth System Grid Federation (ESGF) : <https://www.climate4impact.eu/c4i-frontend/>

Outils et services à la donnée

Description générale du produit de recherche

Nom	Outils et services à la donnée
Description	<p>Le projet ciblé (PC) 2, ou INVEST, du programme de recherche TRACCS, vise également à proposer des services à la données aux autres PC TRACCS afin d'augmenter la valeur ajoutée des données de base par la création de produits de niveaux de complexité supérieurs, pour des calculs simples à des calculs plus complexes nécessitant de grands volumes de données. Ces services viseront à :</p> <ul style="list-style-type: none"> • Accompagner la standardisation de la donnée (e.g., service de contrôle qualité des métadonnées) ; • Visualiser la donnée sous différentes formes (e.g., cartes, graphiques); • Télécharger la donnée (e.g., tout ou partie d'une série temporelle dans différents format); • Effectuer des post-traitements usuels ou complexes à la volée tels que : <ul style="list-style-type: none"> ◦ Calculer des moyennes, ◦ Effectuer une interpolation sur une grille cible, ◦ Calculer des indicateurs climatiques, ◦ Corriger des biais, ◦ Appliquer une descente d'échelle. <p>De nouveaux formats de données plus efficaces permettant des traitements en parallèle et asynchrone sur différents sous-ensemble logiques de données seront également nécessaires (e.g., le format Zarr). Ainsi des services de "conversion" à la volée, avec des systèmes de "caching" seront requis pour permettre ce type de post-traitements gourmand en temps de calcul.</p>
Type	Logiciel
Workpackage	WP3
Mots clés (texte libre)	
Langue	anglais
Date de publication	
Contient des données personnelles ?	Non
Contient des données sensibles ?	Non
Prend en compte des aspects éthiques ?	Non
<hr/>	
Est-ce que des données existantes seront réutilisées ?	
Justification	Les services que proposera le PC2 reposeront en partie sur des outils ou des bibliothèques logicielles existants dans la communauté de modélisation du climat et éprouvés.
Données réutilisées	<ul style="list-style-type: none"> • icclim : https://icclim.readthedocs.io/en/stable/ • xsbck : https://github.com/yrobink/XSBCK • ESMValTool : https://esmvaltool.org/
<hr/>	
Comment seront produites/collectées les nouvelles données ?	

Nom de la méthode**Description**

Un catalogue des services existants sur les trois plateaux techniques partenaires du projet sera établi afin d'établir une spécification, un prototypage et mise en œuvre de services génériques interopérables entre les trois plateformes. La coordination avec les services européens tels que EOSC (European Open Science Cloud), Copernicus ou DestinE (Destination Earth) permettront de s'appuyer sur d'autres API ou interfaces existantes actuellement en cours de développement.

Nature des données

Modèle

Documentation et qualité des données

Simulations et indicateurs climatiques

Quelles métadonnées et quelle documentation (par exemple mode d'organisation des données) accompagneront les données ?

Description

La compréhensibilité des données collectées est assurée par le biais de leurs métadonnées, et en particulier le respect du principal standard de la communauté de modélisation du climat : la convention « [Climate and Forecast \(CF\)](#) » et son extension « [Attribute Conventions Dataset Discovery \(ACDD\)](#) ». Ce standard définit les grands principes de la structure des fichiers ainsi que les attributs clés qu'ils doivent inclure afin d'être "auto-descriptifs". Par conséquent, le premier pré-requis consiste à utiliser des formats de données permettant de stocker les métadonnées associées aux données (c'est le cas de HDF5/NetCDF).

Les données collectées seront ensuite organisées et nommées en respectant les syntaxes de référence (ou *Data Reference Syntax - DRS*) qui définissent des attributs/clés supplémentaires, les noms de fichiers ainsi que les structures de répertoires. Ces exigences s'ajoutent à la convention CF et sont définis pour chaque type de simulation suivant des *vocabulaires contrôlés*.

Chaque plateforme mettra également à jour sa documentation utilisateur en complétant son catalogue de données TRACCS dont il assurera la distribution et l'analyse. Cette documentation reposera sur les bonnes pratiques déjà en vigueur dans la communauté à savoir :

- Lier la documentation aux fichiers ou aux jeux de données (typiquement via l'intermédiaire d'un URL ou "landing page").
- Etablir une documentation minimum pour chaque modèle climatique dont les données collectées sont extraites,
- Employer un vocabulaire contrôlé commun (une initiative est actuellement en cours au niveau WCRP/CMIP pour recenser, servir et aligner les termes sur des concepts ontologiques connues en utilisant les standards de la sémantique web).

Cette documentation pourra s'appuyer sur le projet [ES-DOC](#) (Earth System Documentation) qui vise à développer un écosystème d'outils et de services pour soutenir la création, l'analyse et la diffusion de la documentation sur le système terrestre.

Références associées

- CMIP6 Guidelines : <https://pcmdi.llnl.gov/CMIP6/Guide/modelers.html>
- CORDEX Archive Design : https://is-enes-data.github.io/cordex_archive_specifications.pdf
- CORDEX-Adjust Archive Design : https://is-enes-data.github.io/CORDEX_adjust_drs.pdf
- CMIP5-Adjust Data Reference Syntax : <https://eartharxiv.org/repository/view/2013/>
- CMIP5 Data Reference Syntax and Controlled Vocabularies : https://pcmdi.llnl.gov/mips/cmip5/docs/cmip5_data_reference_syntax.pdf?id=83

Standards de métadonnées/données

- CF (Climate and Forecast) Metadata Conventions : <https://rdamsc.bath.ac.uk/api2/m5>
- Ontologie Schema.org : <https://rdamsc.bath.ac.uk/api2/m101>
- Attribute Convention for Data Discovery 1.3 : https://wiki.esipfed.org/Attribute_Convention_for_Data_Discovery_1-3

Code langue des métadonnées

anglais

Logiciel de documentation

CF Compliance Checker, cf-checker, CMOR (Climate Model Output Rewriter), IOOS Compliance Checker

Quelles seront les méthodes utilisées pour assurer la qualité scientifique des données ?

Description

Un contrôle qualité des données telles que les projections climatiques est une activité scientifique en elle-même et ne repose sur le seul PC2 (e.g., détection des valeurs aberrantes ou de pattern récurrents dans les variables). La validation scientifique des données reposera donc sur une confiance mutuelle tacite avec les partenaires scientifiques des autres projets ciblés de TRACCS. Les jeux de données collectés par le projet INVEST pourront s'accompagner de rapport scientifiques émis par les producteurs de données dans les autres projets ciblés. Les critères d'émission et la structure de ces rapports sera défini en étroite collaboration avec le comité scientifique du projet INVEST. Des services d'errata (comme le [service Errata de ESGF](#)) assureront la documentation transparente des erreurs détectées sur les jeux données ainsi que le suivi des différentes versions et leur généalogie.

Chaque partenaire des PC TRACCS produisant des données devra renseigner les métadonnées du jeu de données en suivant les normes et conventions décrites en 2.1. Le projet INVEST garantit l'intégrité, la qualité et la conformité des métadonnées dans le respect des schémas fixés par les standards choisis. Par conséquent, seules les métadonnées feront l'objet d'un contrôle de qualité systématique à l'aide d'outils communautaires éprouvés et libres avant que les données soient collectées puis distribuées. Une chaîne de traitement appliquera une variété de contrôles des métadonnées et sera notamment mise en place à l'aide des outils suivant :

- "CF-checker" : utilisé pour s'assurer du respect des fichiers netCDF avec la convention CF ;
- "IOOS Checker" : utilisé pour s'assurer de la conformité des métadonnées avec les exigences CMIP et CORDEX ;
- "CMOR" (Climate Model Output Rewriter) : utilisé pour réécrire des fichiers natif de simulations climatique dans les standards CMIP ou CORDEX.
- "nctime" : utilisé pour contrôlé la couverture temporelle d'un fichier et son axe des temps.

Cette chaîne de traitement sera automatisable et configurable en fonction de chaque plateformes. En cas de non conformité avec les standards choisis, un accompagnement à la standardisation des données pourra être apporté par les collecteurs de données du projet INVEST.

Outils et services à la donnée

Quelles métadonnées et quelle documentation (par exemple mode d'organisation des données) accompagneront les données ?

Description

Les produits de recherche correspondant ici à des logiciels ou bibliothèques logicielles ils ne disposent pas de standards de métadonnées. Cependant, le PC2 INVEST s'engage à s'assurer que les codes utilisés ou développés dispose d'une documentation technique et utilisateurs suffisante et à jour, à défaut il y contribuera. Les données produites comporteront des standards déjà existants qui seront étendus pour décrire les nouveaux types de données produites.

Dans le cas des codes scientifiques collectés depuis les partenaires des autres PC TRACCS, il s'agira à minima de décrire la version et de mettre un lien vers la documentation. Cette documentation devra inclure la description des fonctions principales du code et les références vers des notices ou des publications associées.

Quelles seront les méthodes utilisées pour assurer la qualité scientifique des données ?

Description

Bien qu'il n'y ait pas de méthode particulière pour assurer la qualité d'un code ou d'un logiciel, les ingénieurs contribuant au projet partagent leurs expertises en génie logiciel et en science des données. Les expériences de chaque membre seront régulièrement partagées au sein d'INVEST afin d'améliorer l'infrastructure en place et les technologies sur lesquelles elle repose. Des réunions régulières sont organisées au sein des équipes pour discuter de divers aspects et partager l'expertise et les meilleures pratiques.

Exigences légales et éthiques, code de conduite

Simulations et indicateurs climatiques

Comment les autres questions juridiques, comme la titularité ou les droits de propriété intellectuelle sur les données, seront-elles abordées ? Quelle est la législation applicable en la matière ?

Description

Sous l'égide de ses autorités institutionnelles, INVEST s'engage à promouvoir la science ouverte et le libre accès aux ressources et aux publications suivant les principes de l'Open Data. En conséquence, les données réutilisées et collectées au cours du projet sont publiques, libres, ne sont soumises à aucune réglementation en matière de droits d'auteur, de brevets, de marques ou de secrets commerciaux.

Cependant, les textes et éléments graphiques, leur présentation et leur assemblage figurant dans les sites DRIAS, ESPRI et Climate4Impact, et d'une manière générale tous les documents contenus dans ces sites, sont la propriété exclusive du propriétaire du site et ses partenaires ou sont l'objet d'un accord de licence. Ils constituent des œuvres de l'esprit ou des éléments de bases de données et sont protégés en tant que tels par les lois des pays du monde entier et les traités internationaux.

En accédant à ces sites ou à l'un de ses éléments, l'utilisateur accepte une licence concédée aux conditions définies ci-après.

Article 1er : droit d'usage

Au titre de la présente licence, est exclusivement concédé un droit d'usage privé, individuel, personnel et non transmissible sur le contenu du site ou de l'un de ses éléments.

Article 2 : réutilisation

Les textes, éléments graphiques et documents contenus dans ces sites peuvent être réutilisés sous réserve qu'ils ne soient pas altérés, que leur sens ne soit pas dénaturé et que leur source et la date de leur dernière mise à jour soient mentionnées. Leur redistribution en l'état n'est pas autorisée.

- Les «partenaires» se réservent le droit de modifier ces conditions à tout moment
- Les «partenaires» se réservent le droit de modifier ou de corriger le contenu de ce site à tout moment et sans préavis, notamment les textes, éléments graphiques et documents contenus dans ce site.
- Tous les droits qui ne sont pas expressément concédés aux termes des présentes sont réservés par les «partenaires».

Les «partenaires», ni l'Etat, ne pourront être tenus responsables de dommages résultant de l'interprétation, de l'utilisation, d'une impossibilité d'utilisation ou d'une impossibilité d'accès aux informations disponibles sur ces portail, ainsi que d'une perte éventuelle provenant de cette utilisation ou de cette impossibilité.

Comment les éventuelles questions éthiques seront-elles prises en compte, les codes déontologiques respectés ?

Description

Le projet INVEST collectera des données non sensibles. Par conséquent, les normes disciplinaires légales et éthiques ne s'appliquent pas (ex. RGPD).

Outils et services à la donnée

Comment les autres questions juridiques, comme la titularité ou les droits de propriété intellectuelle sur les données, seront-elles abordées ? Quelle est la législation applicable en la matière ?

Description

Le PC2 INVEST suivra la politique et les recommandations "Open Science" activement promues par nos institutions nationales et nos partenaires européens. En conséquence, les codes et logiciels réutilisés et collectés au cours du projet sont publiques, libres, ne sont soumis à aucune réglementation en matière de droits d'auteur, de brevets, de marques ou de secrets commerciaux.

Cependant, les textes et éléments graphiques, leur présentation et leur assemblage figurant dans les sites DRIAS, ESPRI et Climate4Impact, et d'une manière générale tous les documents contenus dans ces sites, sont la propriété exclusive du propriétaire du site et ses partenaires ou sont l'objet d'un accord de licence. Ils constituent des œuvres de l'esprit ou des éléments de bases de données et sont protégés en tant que tels par les lois des pays du monde entier et les traités internationaux.

En accédant à ces sites ou à l'un de ses éléments, l'utilisateur accepte une licence concédée aux conditions définies ci-après.

Article 1er : droit d'usage

Au titre de la présente licence, est exclusivement concédé un droit d'usage privé, individuel, personnel et non transmissible sur le contenu du site ou de l'un de ses éléments.

Article 2 : réutilisation

Les textes, éléments graphiques et documents contenus dans ces sites peuvent être réutilisés sous réserve qu'ils ne soient pas altérés, que leur sens ne soit pas dénaturé et que leur source et la date de leur dernière mise à jour soient mentionnées. Leur redistribution en l'état n'est pas autorisée.

- Les «partenaires» se réservent le droit de modifier ces conditions à tout moment
- Les «partenaires» se réservent le droit de modifier ou de corriger le contenu de ce site à tout moment et sans préavis, notamment les textes, éléments graphiques et documents contenus dans ce site.
- Tous les droits qui ne sont pas expressément concédés aux termes des présentes sont réservés par les «partenaires».

Les «partenaires», ni l'Etat, ne pourront être tenus responsables de dommages résultant de l'interprétation, de l'utilisation, d'une impossibilité d'utilisation ou d'une impossibilité d'accès aux informations disponibles sur ces portail, ainsi que d'une perte éventuelle provenant de cette utilisation ou de cette impossibilité.

Comment les éventuelles questions éthiques seront-elles prises en compte, les codes déontologiques respectés ?

Description

Le projet INVEST collectera des outils non sensibles. Par conséquent, les normes disciplinaires légales et éthiques ne s'appliquent pas (ex. RGPD).

Traitement et analyse des données

Simulations et indicateurs climatiques

Comment et avec quels moyens seront traitées les données ?

Description

Le projet INVEST vise à fournir des services à la donnée haut de gamme pour les utilisateurs avancés qui s'appuient sur les développements réalisés dans le programme TRACCS.

Dans un premier temps, les services actuellement déployés sur les 3 plateaux techniques qui sous-tendent le projet seront étendus aux données collectés par le projet INVEST :

- La plateforme DRIAS offre plusieurs espaces d'appropriation des données : un espace éditorial, un espace de visualisation interactive des jeux de données, enfin un espace dédié au téléchargement (espace Données&Produits) permet d'extraire un sous-ensemble de données (géographique ou temporelle).
- La plateforme Climate4Impact offre des espaces de travail dédiés au calcul d'indicateurs à la volée reposant sur la bibliothèque Python "iclim". Son interface web permet également d'extraire un sous-ensemble de données (géographique ou temporelle).
- La plateforme ESPRI offre des espaces de travail ou de projet avec un accès à différents environnements virtuels de recherche (ensemble de bibliothèques logicielles dédié à l'analyse de données issues de simulations climatiques). Un service de traitement à la demande et à distance permet de récupérer également de récupérer une extraction géographique ou temporelle d'un jeu de données mais également de produire à la volée une moyenne (journalière, mensuelle, saisonnière) ainsi qu'une interpolation sur une grille spatiale régulière.

Lors de la phase de développement du projet, une couche de préparation des données sera ajoutée pour un traitement efficace, basé sur des technologies et des formats de fichier "cloud-ready" (Zarr et Dask). Les différents protocoles de traitement à distance des plateformes seront également complétés des méthodes statistiques innovantes qui seront développés dans le cadre des autres projets TRACCS, comme de nouveaux algorithmes de descentes d'échelle ou de correction de biais.

Equipements, plateaux techniques

- Mésocentre ESPRI : <https://cat.opidor.fr/index.php/Mésocentre ESPRI>

Outils et services à la donnée

Comment et avec quels moyens seront traitées les données ?

Description

Les moyens propres aux plateformes techniques mobilisées pour le projet seront mises en œuvre pour la gestion et la documentation des codes et logiciels utilisés ou développés.

Equipements, plateaux techniques

- Mésocentre ESPRI : <https://cat.opidor.fr/index.php/Mésocentre ESPRI>

Stockage et sauvegarde des données pendant le processus de recherche

Simulations et indicateurs climatiques

Comment les données seront-elles stockées et sauvegardées tout au long du projet ?

Besoins de stockage

En fonction de la nature et de l'usage des différents jeux de données produits par les autres PC du programme TRACCS, il sera nécessaire en premier lieu de déterminer le périmètre et la plateforme au sein de laquelle il est pertinent de les collecter.

Un aspect important associé au stockage est la conception et la mise en place d'un catalogue des données gérées par INVEST.

Ce catalogue devra être en capacité :

- de décrire aussi bien des jeux de données initiaux, des données post-traitées (comme les indicateurs climatiques) et des services
- de décrire des espaces de stockage distribués : celui d'ESPRI et celui de DRIAS notamment

Il est proposé que les projections climatiques régionales et globales standardisées soient stockées sur la plateforme ESPRI qui dispose de toutes les procédures et de l'expertise nécessaire à une gestion FAIR de ce type de données. ESPRI gère sa propre infrastructure de stockage répartie sur deux sites (Ecole Polytechnique et Sorbonne Université). Il bénéficie également d'un stockage dédié hébergé et géré par deux centres nationaux de calcul intensif (IDRIS et TGCC). Le stockage des données primaires se fait principalement sur disque. L'infrastructure comprend également des bibliothèques de bandes magnétiques utilisées principalement pour la sauvegarde et l'archivage des données. L'infrastructure de stockage de l'ESPRI repose sur des équipements standard, éprouvés et testés, constitués de baies de stockage dont les volumes logiques sont sécurisés par des systèmes de "pool" (évolution des systèmes RAID). Deux serveurs de fichiers assurent un accès redondant et contrôlent chaque baie. Le stockage de l'ESPRI utilise un système de fichiers LUSTRE, qui permet d'agréger les espaces disques en volumes logiques unifiés et garantit des performances d'accès très élevées et une extension du stockage sans perte d'efficacité. La gestion des risques est également assurée par les ingénieurs d'ESPRI grâce à une surveillance continue des ressources de stockage sur site : une supervision de bas niveau des services (état des contrôleurs, des alimentations, des disques logiques, physiques et virtuels, des ventilateurs, de la température, etc.) ainsi qu'une supervision des services de haut niveau sont effectuées par des sondes Nagios et avertissent en temps réel en cas d'alerte critique due à une défaillance du système.

Les jeux de données post-traités, comme des indicateurs climatiques pré-calculés sur les territoires nationaux (métropole et outre-mer), seront stockés sur un espace de stockage associé au portail DRIAS.

A ce stade du projet, l'estimation est difficile d'estimer la volumétrie nécessaire au projet INVEST, mais elle pourrait atteindre quelques PetaOctets.

Volume estimé des données

1

Unité

Po

Equipements, plateaux techniques

- Mésocentre ESPRI : <https://cat.opidor.fr/index.php/Mésocentre ESPRI>

Outils et services à la donnée

Comment les données seront-elles stockées et sauvegardées tout au long du projet ?

Besoins de stockage

Le PC2 INVEST s'engage également à suivre les bonnes pratiques en matière de gestion et de pérennisation des codes, à savoir :

- Reprendre ou héberger les codes sur des dépôts Git publics (i.e., sans restriction d'accès) et permettant le versionnement des outils.
- S'assurer que les codes utilisés sont maintenus par une communauté robuste,
- Poser des DOI sur les codes ou les publications qui les accompagnent.

Dans le cas où des outils devront être développés par le PC2, ces mêmes recommandations seront appliquées.

Equipements, plateaux techniques

- CC-IN2P3 : <https://cat.opidor.fr/index.php/CC-IN2P3>

Partage des données et conservation à long terme

Simulations et indicateurs climatiques

Comment les données seront-elles partagées ?

Modalités de partage

Un principe de base de l'Open Data est de distribuer les données publiquement pour qu'elles soient ensuite analysées par la communauté scientifique. Conformément à ce principe, le projet INVEST encourage la mise à disposition des données aussi rapidement que nécessaire pour préserver la valeur des données et au plus grand nombre d'utilisateurs pour le plus grand nombre d'objectifs. Aucun risque de divulgation sensible n'ayant été identifié, aucune période d'embargo ne sera appliquée sur les données partagées par le projet INVEST.

La réutilisation des données est un autre pilier de l'Open Science. A cette fin, le projet INVEST applique autant que possible les principes FAIR dans la distribution des données collectées au sein de son infrastructure partagée. Les procédures d'accès aux données dépendent du type de données et des utilisateurs. D'une manière générale, l'ESPRI offre un accès pour la découverte et le téléchargement des données via des interfaces web (catalogue GeoNetwork, serveurs THREDDS, portails web spécifiques) et en utilisant différents protocoles (HTTPS, SFTP, GridFTP, WPS). Les utilisateurs ayant accès aux ressources du centre de calcul ESPRI peuvent bénéficier d'un accès direct au système de fichiers hébergeant les arborescences de données. Le portail Climate4Impact constitue une porte d'entrée plus accessible à un catalogue de donnée identique mais autant pour des utilisateurs scientifiques (sciences du climat et autres sciences également) et non-scientifiques.

L'accès aux données DRIAS se fait aujourd'hui uniquement via une interface web, qui offre toutefois plusieurs modalités d'accès aux données :

- une sélection paramétrable sur toutes les caractéristiques de l'information : spatial, temporel, modèle, RCP/SSP, variable, choix du format de restitution (csv/netcdf), ...
- une sélection plus globale par modèle, RCP et variable en format netcdf uniquement

Dans l'avenir un accès de type webservices sera proposé pour accéder aussi bien aux jeux de données corrigées qu'aux données post-traitées de type indicateur climatique.

Potentiel de réutilisation Réutilisation totale.

Entrepôt/Catalogue de données

- ESPRI : <https://cat.opidor.fr/index.php/ESPRI> (CoreTrustSeal)

Comment les données seront-elles conservées à long terme ?

Justification

Par défaut, les plateformes du projet INVEST s'engagent à conserver les données collectées au moins sur la durée du projet.

Au delà de cette période, les données ne sont pas strictement archivées, mais elles sont préservées par la garantie appliquée au matériel de stockage dédié qui est renouvelée tous les 7 ans. Les versions anciennes/dépréciées sont temporairement conservées pendant une période suffisamment longue afin de garantir la reproductibilité des analyses récentes de TRACCS, leur durée de conservation s'étend généralement à quelques années après l'arrivée d'une nouvelle version du simulation donnée.

Pour les données non reproductibles et immuables dont les plateformes du projet INVEST seront les premiers dépositaires, la durée de conservation peut être bien plus longue voire illimitée.

Volume estimé des données

2

Unité

Po

Date de début

Date de fin

Archive

Titre	ESPRI - Centre de Calcul et de Données de l'IPSL
Identifiant de l'entrepôt	https://espri.ipsl.fr/
Type d'identifiant	URL
Politique de données	https://cloud.ipsl.fr/index.php/s/CgkHcy6nfBongEM
Localisation géographique (pays)	FR
Certification	CoreTrustSeal
Système d'identifiants	DOI
Politique de gestion des versions	
Standards de métadonnées disponibles	CF (Climate and Forecast) Metadata Conventions

Dispositions finales

Outils et services à la donnée

Comment les données seront-elles partagées ?

Modalités de partage

Les codes et logiciels utilisés au cours du projet INVEST sont principalement écrit en langage Python. Ce langage syntaxique permet d'adresser plusieurs aspect d'un même outil (API, réseau, CLI) et une documentation technique auto-générée. Il est largement répandu dans la communauté.

Le dépôt des codes et logiciels sous Git permettra un partage sans restriction d'accès et une totale transparence sur les fonctionnalités des services INVEST. Chaque version des codes seront disponibles sans délais bien que généralement seules les versions finales seront documentées.

Potentiel de réutilisation

Comment les données seront-elles conservées à long terme ?

Justification

Les codes logiciels sont gérés tout au long de leur cycle de vie, du développement à la distribution et au débogage, sur une plateforme Git qui compte-tenu du faible coût en stockage permettra une conservation sans limites des codes.

Si un code ne dispose pas déjà de son dépôt officiel, ESPRI recommande d'utiliser la plateforme Gitlab hébergée par le Centre de Calcul de l'Institut National de Physique Nucléaire et de Physique des Particules (CC-IN2P3).

Date de début**Date de fin****Dispositions finales**