Charte d'accès à la plateforme Biochem-Env

Préambule

Dans le cadre du Programme Investissements d'Avenir « Infrastructures Nationales de Biologie et Santé » relatif à l'action « Santé et Biotechnologie » lancé en 2011 par le Commissariat Général à l'Investissement (CGI), le Ministère de l'Enseignement Supérieur et de la Recherche et l'Agence Nationale de la Recherche (ANR), le CNRS, INRAE et l'UGA se sont associés pour déposer un projet intitulé « Analyse et Expérimentation sur les Ecosystèmes – France (AnaEE-France) ».

La présente charte a pour objectif de définir les modalités d'accès à la plateforme de recherche Biochem-Env, membre de l'Infrastructure de Recherche AnaEE-France, et de fixer les droits et obligations des utilisateurs.

S'informer sur la plateforme et la contacter

Les utilisateurs peuvent s'informer sur la plateforme et ses activités via :

i) son portail internet : http://www.biochemenv.fr

ii) sa page Plug In Labs : https://www.pluginlabs-universiteparissaclay.fr/en/fiche/ecosys-plateforme-biochem-env-2/

iii) son compte LinkedIn : Biochem-Env

iv) sa chaine Youtube : Biochem-Env

v) sa publication de présentation : https://doi.org/10.1007/s11356-017-8973-x

Biochemenv, c'est également :

vi) une collection Hal : https://hal.inrae.fr/BIOCHEMENV/

vii) un entrepôt de données : https://entrepot.recherche.data.gouv.fr/dataverse/biochem-env

viii) un DOI : https://doi.org/10.15454/HA6V6Y

ix) un Plan de gestion de Données : https://hal.inrae.fr/hal-05171437v1

Les utilisateurs peuvent contacter la plateforme via l'adresse : contact-biochemenv@inrae.fr

Les utilisateurs peuvent soumettre leurs demandes de projets via le portail : https://isia.cnrs.fr/

Open licence - etalab

Article 1 - Objet de la charte

Cette charte présente :

- Le dispositif et l'offre de service,
- La gouvernance de la plateforme,
- La procédure de soumission des projets,
- La procédure de sélection des projets,
- Les modalités d'utilisation de la plateforme et les obligations associées,
- Le cadre conventionnel de l'accueil en vue de l'utilisation de la plateforme.

Article 2 - Présentation de la plateforme et de l'offre de service

Contexte

Créée en 2012, la plateforme Biochem-Env est située sur le Campus Agro Paris-Saclay de Palaiseau au sein de l'UMR 1402 ECOSYS (centre INRAE Ile-de-France Versailles-Saclay).

Biochem-Env est une plateforme de recherche qui permet le développement et la mesure d'indicateurs biochimiques dans l'environnement et les organismes des écosystèmes continentaux.

Les missions de la plateforme sont les suivantes :

i) la **conduite de projets de recherche** et la **réalisation de prestations** portant sur l'observation et l'expérimentation sur les écosystèmes (prélèvement des échantillons, analyses, interprétation des résultats…)

ii) le **développement et** le **transfert des méthodes** de mesure et d'**interprétation des résultats**, incluant la **normalisation**,

iii) la **mise à disposition des compétences, matériels et locaux** pour l'analyse (accueil et accompagnement des utilisateurs)

iv) la réalisation d'actions d'**expertise**, de **veille** scientifique et technique, ainsi que de **formation,** notamment de doctorants.

Offre de service

Dans l'environnement (sols et sédiments), la plateforme permet actuellement la mesure d'activités enzymatiques impliquées dans les cycles biogéochimiques et le métabolisme des macromolécules, et/ou traduisant l'activité métabolique globale de ces écosystèmes. Elle réalise également la mesure d'indicateurs biochimiques chez les organismes (invertébrés benthiques et terrestres, plantes) comme les réserves énergétiques et macromolécules, les enzymes du stress oxydant et/ou associées aux mécanismes de détoxication, l'exposition aux contaminants environnementaux…).

Le catalogue des indicateurs actuellement mesurés est disponible sur le portail internet de la plateforme.

Fonctionnement de la plateforme et Qualité

La plateforme Biochem-Env est engagée depuis sa création dans une démarche Qualité. Son Système de Management de la Qualité, basé sur l'approche processus et l'analyse des facteurs de succès (approche par les risques et les opportunités), s'appuie actuellement sur les principes de l'amélioration continue de son fonctionnement et de la satisfaction de ses parties intéressées pertinentes. La déclaration de politique générale de la plateforme est disponible sur son portail internet. La plateforme est labellisée IBISA et certifié ISO9001:2015 depuis novembre 2017. Elle est également labellisée par le GIS IBiSA et INRAE en tant qu'Infrastructure Scientifique Collective (ISC°

Ainsi, la plateforme dispose **d'un management assurant** :

- la maîtrise de ses activités scientifiques et techniques, en termes de fiabilité des résultats d'analyse et de **traçabilité des expérimentations**, notamment dans le domaine de la conception et du développement des méthodes**)**,

- **la maîtrise des échantillons, des données et de leur traçabilité**, au moyen d'un LIMS (opérationnel en 2018) interopérable avec d'autres plateformes analytiques INRAE, afin d'avoir un suivi global de l'ensemble des échantillons gérés par la plateforme,

- **la maîtrise de ses matériels** par un suivi métrologique et réglementaire adapté et régulier, encadré par la procédure de gestion du parc d'équipements. Le SMQ permet ainsi à la plateforme de connaître l'état d'opérationnalité de tous ses équipements, autant ceux du laboratoire que de terrain.

- la maîtrise de sa documentation et de son système d'information grâce à l'utilisation d'un espace collaboratif. Ainsi, la plateforme a rédigé l'ensemble des processus, procédures, modes opératoires, instructions, enregistrements nécessaires à son fonctionnement et gère leur mise à jour.

La plateforme réalise ses activités en accord avec les grands principes de la charte des infrastructures d'INRAE : ouverture, partage, transparence, accès aux données, soutenabilité financière, conformité, ainsi qu'avec les orientations stratégiques de l'institut décrit dans son document d'orientation INRAE 2030.

Article 3 - Gouvernance de la plateforme

Le Service est géré par un Responsable Scientifique et une Responsable Opérationnelle qui animent un Comité Local de Plateforme. Ce comité comprend :

- Les responsables de la plateforme,
- Les personnels impliqués dans le fonctionnement de la plateforme,
- Des scientifiques invités.

Le Comité Local a pour missions le suivi au quotidien des plateformes et la sélection des demandes soumises. Le CLP peut être élargi à une personnalité extérieure issue du Comité d'Orientation Scientifique (COS) de la plateforme pour tout arbitrage rendu nécessaire, notamment pour la sélection des projets.

Article 4 - Procédure de soumission des projets

Tout demandeur (utilisateur potentiel) doit contacter les responsables de la plateforme via contact-biochemenv@inrae.fr afin de préciser les grandes lignes de sa demande ainsi que les modalités de soumission. A l'issue des premiers échanges, et dans le cas de projets de recherche en partenariat, le demandeur s'inscrit sur l'interface ISIA (https://isia.cnrs.fr/) sur laquelle il enregistre se demande dans un délai de deux semaines. L'interface est également disponible depuis le site AnaEE-France (http://www.anaee-france.fr/fr/) ou le site Biochem-Env (http://www.biochemenv.fr/).

Article 5 - Procédure de sélection des demandes

5.1 Sélections des demandes

Les demandes de partenariats ou de prestations soumises sur l'interface ISIA ou adressés par mail à la plateforme sont examinés par le Comité Local de Plateforme, au fil de l'eau lors de ses réunions hebdomadaires, et donnent lieu à l'établissement d'une fiche de faisabilité.

- Les demandes co-construites par des partenaires académiques avec la direction de la plateforme (projets en partenariats de recherche) sont discutés en Comité Local de Plateforme pour préciser le calendrier de programmation de leurs différentes étapes.
- Les demandes qui n'ont pas été co-construites par les partenaires académiques avec la direction de la plateforme (prestations d'analyses dans le cas de projets évalués par des pairs), sont dispensés d'une évaluation scientifique par le Comité Local de Plateforme. Celui-ci se donne toutefois le droit de se prononcer sur leur faisabilité technique (re-définition des méthodes et re-dimensionnement du nombre d'échantillons) et le calendrier de programmation.
- Les demandes qui n'ont pas été co-construites (le plus souvent prestations d'analyses issues de demandeurs privés), et qui n'ont pas fait l'objet d'une évaluation préalable par des pairs, sont évaluées par le Comité Local de Plateforme sur le plan scientifique ainsi que sur leur faisabilité technique et le calendrier de programmation.

Après relecture et si besoin rédaction d'une fiche de faisabilité, le Comité Local émettra un avis définitif sur la demande :

- Acceptée,
- Ou acceptée sous conditions
- Ou rejetée.

Les informations finales seront transmises à chaque demandeur ; et discutées à leurs demandes.

5.2 Critères de sélection

Les critères de sélection des projets proposés, définis par le Comité Local de Plateforme, concernent : l'intérêt scientifique de la proposition au regard des missions de la plateforme, sa faisabilité technique, et la charge supportée par la plateforme (disponibilité de personnel, matériels, locaux).

Si les demandes recevables sont supérieures à la capacité d'accueil de la plateforme, les critères suivants seront considérés pour établir des priorités :

- Partenariats de recherche dans lesquels la plateforme assure le rôle de porteur de projet ou de pilote d'un axe de travail (work package),
- Partenariats et prestations de recherche avec les secteurs académiques ou privés, pour lesquels la plateforme est partenaire sans être porteuse du projet ni pilote d'un axe de travail,
- Autres projets.

La demande pourra être jugée non recevable en cas de conflit d'intérêts ou de problème éthique. En cas de litige, le Comité de Direction d'AnaEE-France sera saisie pour donner une décision finale.

5.3 Suivi de la sélection

Via ISIA, le Comité de Direction d'AnaEE-France a une vision globale des projets menés au sein de l'Infrastructure et est garant de l'équité de sélection des projets. Le Conseil Scientifique d'AnaEE-France peut questionner la plateforme concernant l'adéquation des projets avec les objectifs de l'Infrastructure et l'ouverture internationale.

Article 6 - Modalités d'utilisation de la plateforme et obligations associées

6.1 Modalités d'utilisation

6.1.1 Catégories d'utilisateurs

Les utilisateurs de la plateforme se répartissent en 3 catégories :

- Les chercheurs, enseignants-chercheurs et personnels d'unités sous tutelle INRAE ;
- Les chercheurs, enseignants-chercheurs et personnels de la recherche du secteur académique ;
- Les acteurs du secteur privé (sociétés, bureaux d'étude…) ou de la société civile (associations).

Des règles de tarification différentes s'appliquent à chacune de ces catégories, la troisième étant soumise aux coûts complets.

6.1.2 Accès aux locaux

Tout accès aux locaux doit faire l'objet d'une demande et s'inscrire dans un cadre conventionné selon les procédures en vigueur dans l'UMR ECOSYS (stage, accueil visiteurs, etc.). Une procédure particulière s'applique au laboratoire de quarantaine de la plateforme.

6.1.3. Accès aux équipements

L'accès aux équipements doit également faire l'objet d'une demande et doit être planifié avec les responsables des équipements identifiés.

6.1.4. Accès aux échantillons

Les échantillons fournis à la plateforme par ses utilisateurs, ou prélevés par la plateforme pour leur compte, sont sous la responsabilité des utilisateurs.

A l'issue des analyses, la plateforme questionne les utilisateurs sur le devenir de leurs échantillons : destruction, restitution à leurs frais ou session à la plateforme. Certains échantillons pourront en effet être conservés par la plateforme pour leur intérêt ou à des fins de développement et d'amélioration de ses méthodes, analyses ou expérimentations, sous réserve de l'accord formel des utilisateurs.

Les utilisateurs de la plateforme ont accès aux échantillons qu'ils ont fournis et uniquement à ceux-ci.

Une attention particulière est apportée aux matériels de quarantaine qui nécessitent auprès de la plateforme des démarches règlementées pour leur introduction, manipulation, détention et élimination.

6.1.5. Accès aux données

Les données sont soumises aux règlementations françaises (république numérique) et européennes relative à la Science Ouverte.

L'ensemble des données et métadonnées issues des plateformes AnaEE-France achevées et financées majoritairement sur des fonds publics sont par conséquent communicables librement et sans contrepartie financière, autre que la part des frais spécifiques qui pourraient être engagés pour répondre aux demandes, à toute personne qui en fait la demande, sauf exceptions légales détaillées dans le guide d'analyse du cadre juridique des données (par exemple lorsque la protection de données personnelles est en jeu, ou pour des raisons de confidentialité sécurité défense ou pour des raisons légitimes invoquées par l'utilisateur concerné). La responsabilité de la communicabilité de la donnée incombe à son propriétaire. La politique générale des données AnaEE-France est présentée en Annexe 1.

6.1.5.1. Fourniture des données produites par la plateforme

Les données sont ouvertes selon la politique "ouvert autant que possible, fermé autant que nécessaire". Selon la nature des échantillons et des conventions de prestations, ceci peut se faire en accès public, en accès privé ou après une période d'embargo.

- La diffusion et le partage des données sont définis dans l'accord de consortium dans le cadre de partenariats de recherche académique.

- La diffusion et le partage des données sont discutés et précisés dans le cadre de prestations avec des partenaires de recherche académique.

- Les données restent confidentielles dans le cas de partenariats ou de prestations financés par des acteurs du secteur privé ou de la société civile. Toutefois, des conventions pour leur partage et leur valorisation peuvent être mises en place.

6.1.5.2 Contribution aux référentiels de la plateforme

Sauf refus explicite des utilisateurs privés ou de la société civile pour lesquels les données doivent rester confidentielles, toutes les données produites par la plateforme et les métadonnées associées fournies par les utilisateurs pour leur interprétation contribuent aux référentiels de la plateforme. Les métadonnées associées aux indicateurs biologiques mesurés par la plateforme sont précisées en Annexe 2.

Les données produites par la plateforme dans le cadre d'une expérience réalisée sur une plateforme expérimentale AnaEE doivent être gérées dans les mêmes conditions que toutes les données acquises

6.1.6. Accompagnement en moyens humains

Le Service s'engage à fournir des moyens humains nécessaires au bon déroulement des partenariats et prestations. Dans le cadre de sa mission de formation, il veillera à l'encadrement des doctorants accueillis (et autres stagiaires) dont la participation à la réalisation des analyses est indispensable.

6.2 Obligations associées à l'utilisation du service

6.2.1 Respect du règlement intérieur de l'unité de gestion du Service

L'utilisateur s'engage à respecter le règlement intérieur de l'UMR ECOSYS ainsi que, si applicable, le règlement relatif au laboratoire de quarantaine de la plateforme (voir 6.2.5).

6.2.2. Respect de l'environnement et expérimentation en milieu naturel

L'unité de gestion de la plateforme met en place les outils nécessaires pour minimiser l'impact de ses activités sur l'environnement et minimiser l'utilisation des ressources : gestion/recyclage des déchets, gestion énergétique des bâtiments, gestion des ressources en eau, etc.

Les travaux réalisés par ou avec la plateforme doivent pouvoir servir de référence en matière de respect de l'environnement (dégradations contrôlées sur l'environnement).

Au champ, la plateforme et ses utilisateurs limitent l'empreinte environnementale des prélèvements réalisés : utilisation des techniques les moins destructrices, limitation de l'utilisation ou remplacement de produits chimiques toxiques, restauration des milieux au plus proche de leur état d'origine.

6.2.3 Droits et obligations liés à l'utilisation des locaux

Le Service s'engage à mettre à disposition les locaux nécessaires à la réalisation de certains projets. L'utilisateur s'engage à les restituer dans leur état initial.

6.2.4 Droits et obligations liés à l'expérimentation animale

Le Service s'engage à suivre la règlementation relative à l'expérimentation animale, si applicable.

6.2.5 Droits et obligations liés à l'introduction, la manipulation et la détention de matériels de quarantaine

La plateforme bénéficie d'un agrément pour l'introduction, la manipulation et la détention de matériels de quarantaine valable jusqu'en février 2028. Ces matériels comprennent : le sol vivant, tamisé, séché ou congelé, végétaux naturels, séchés parfois broyés ou congelés, matériels biologiques frais, congelés ou fixés dans l'alcool, originaires de pays non Européens. L'introduction ou la mise en circulation de tout matériel couvert par cet agrément doit en permanence être accompagné par une Lettre Officielle d'Autorisation délivrée par le ministère.

6.2.6 Droits et obligations liés à l'utilisation des équipements et des consommables

L'utilisation des équipements et des consommables (produits et réactifs) doit faire l'objet au préalable d'une formation des utilisateurs, en termes d'hygiène et sécurité, et de métrologie. Chaque utilisateur s'engage à respecter les règles d'utilisation des équipements et des consommables définies par les responsables du Service. Les frais de remise en état des équipements consécutifs à leur détérioration (mauvaise utilisation) sont à la charge de l'utilisateur.

6.2.7 Valorisation et communication

La plateforme Biochem-Env, l'UMR ECOSYS et l'infrastructure AnaEE-France doivent être associées à toute forme de valorisation et de diffusion des données et résultats ayant mobilisé ses compétences, matériels et locaux, dans le cadre de partenariats et de prestations avec le secteur académique.

Les productions doivent obligatoirement mentionner dans les remerciements l'aide de l'ANR par la phrase suivante :

- Ce travail a bénéficié, au titre de l'utilisation de la plateforme Biochem-Env (DOI 10.15454/HA6V6Y, UMR 1402 ECOSYS, centre INRAE IdF Versailles-Saclay), labellisée Infrastructure Scientifique Collective (ISC), d'une aide de l'Etat gérée par l'Agence Nationale de la Recherche au titre de France 2030 portant la référence "ANR-24-INBS-0001" AnaEE France.».

- This work benefited from the use of the Biochem-Env platform (DOI 10.15454/HA6V6Y, UMR 1402 ECOSYS, INRAE IdF Versailles-Saclay centre), labelled as an ISC (Infrastructure Scientifique Collective), and supported by the French government through the Agence Nationale de la Recherche (ANR) under the France 2030 programme, reference "ANR-24-INBS-0001" AnaEE France.

Les personnels de la plateforme co-auteurs des productions doivent être mentionnés avec la double affiliation :

- Université Paris-Saclay, INRAE, AgroParisTech, UMR EcoSys, Campus Agro, F-91120 Palaiseau, France
- Université Paris-Saclay, INRAE, Plateforme Biochem-Env, F-91120 Palaiseau, France

La plateforme doit être citée dans la partie expérimentale des productions avec un renvoi vers les références bibliographiques selon la phrase :

- Biochem-Env, INRAE, 2019. Plateforme de Biochimie Environnementale, https://doi.org/10.15454/HA6V6Y

Les supports visuels (posters présentations...) doivent comporter les logos de la plateforme Biochem-Env, de l'infrastructure AnaEE-France et France 2030, qui seront fournis sur demande.

Conformément à la loi pour une République numérique de 2016, la plateforme se réserve le droit de valoriser les résultats qu'elle a obtenus dans le cadre de partenariats et prestations avec des utilisateurs académiques, à l'issue d'un délai de 5 ans suivant la réalisation des analyses si le porteur académique n'a proposé aucune forme de valorisation. Avant d'engager toute démarche de valorisation, la plateforme prendra contact avec le porteur de projet afin d'échanger sur les possibilités de valorisation des résultats et de définir conjointement les modalités les plus appropriées pour leur mise en œuvre.

6.2.8 Avertissements dans le cadre de prestations

Dans le cas de prestations (hors contrats partenariaux de recherche), la plateforme s'engage à réaliser les analyses selon la demande formulée et à mettre en œuvre toute la diligence requise, et ceci en fonction des connaissances scientifiques et de l'état de la technique à la date d'exécution desdites prestations.

Le demandeur de la prestation utilisera les résultats obtenus sous sa seule responsabilité à l'égard de tiers éventuels. Il renoncera à tout recours contre la plateforme pour tout dommage qui pourrait résulter de l'utilisation, de l'interprétation, de l'impossibilité d'utilisation ou de l'exploitation commerciale des résultats fournis.

La plateforme ne saurait être tenue responsable de la qualité des résultats lorsque les échantillons sont prélevés et fournis par le demandeur sans concertation sur les protocoles de prélèvement et de transport.

6.2.9 Traitement des données à caractère sensible

Des données à caractère personnel sont collectées par les responsables de la plateforme Biochem-Env (contact-biochemenv@inrae.fr) afin de permettre son fonctionnement interne. La base légale du traitement repose sur notre mission d'intérêt public de production de connaissances. Ces données sont conservées 10 ans après la réception des demandes pour assurer leur gestion et leur suivi, le rendu des résultats et leur valorisation, jusqu'à publication des résultats agrégés pour la réalisation d'enquêtes de satisfaction (partenaires/clients, personnels, fournisseurs), 5 ans à compter de la résolution de réclamations éventuelles.

Tout partenaire/client, personnel, fournisseur de la plateforme dispose de droits d'accès, de rectification, d'opposition, de limitation et d'effacement de ses données personnelles auprès de la plateforme (contact-biochemenv@inrae.fr), ainsi que de la possibilité de solliciter la Déléguée à la protection des données INRAE (cil-dpo@inrae.fr) et d'adresser une réclamation à la CNIL (https://www.cnil.fr/fr) en cas de litige.

Article 7 - Cadre conventionnel de l'accueil en vue de l'utilisation de la plateforme

Tout accès à la plateforme doit être formalisé dans le respect des procédures administratives et financières en vigueur sur le centre INRAE Ile-de-France Versailles-Saclay.

Les éventuels différends rencontrés, s'ils ne peuvent être réglés à l'amiable, seront portés à connaissance du Comité de Direction d'AnaEE-France.

En cas de non-respect des règles définies dans la présente charte, les responsables de la plateforme, sans préjuger des poursuites ou procédures de sanctions pouvant être engagées à l'encontre des utilisateurs par les personnes lésées, peuvent interdire l'accès aux installations et aux données par mesure conservatoire.

Annexe 1 : politique générale des données AnaEE-France

Politique des données AnaEE-France

Ce document définit la politique des données acquises à partir des services de l'infrastructure nationale AnaEE-France et leur modalité de diffusion. Nous rappelons que ces services couvrent l'accès aux plateformes expérimentales, la production d'analyses pour caractériser les écosystèmes, la mise à disposition d'instruments et l'accès aux plateformes de modélisation. Ce document ne traite pas des données générées par des modèles utilisant les plateformes de modélisation qui font l'objet d'un document spécifique.

Ce document constitue un cadre définissant les **principes d'utilisation et de diffusion des données** qui devra trouver par la suite une traduction juridique dans les documents appropriés (licences, chartes) et s'appuiera sur des outils informatiques de mise à disposition des données. Ce document met à jour les principes énoncés dans l'accord de consortium AnaEE-France. Ce document devra évoluer pour permettre un alignement avec la politique qui sera décidée au niveau Européen tout en restant conforme à la législation Française et aux politiques institutionnelles (CNRS, INRA et autres tutelles des plateformes AnaEE-France).

Métadonnées caractérisant les projets d'expérimentation

Toutes les demandes d'utilisation des services font l'objet d'une rédaction de projet définissant la conception de l'expérience et les ressources à mobiliser. Le projet soumis via une interface de soumission de projet permet de collecter des métadonnées qui seront visibles, sans restriction, sur le site AnaEE-FR, dès la fin du projet. Les métadonnées « projet » comportent :

- Titre et acronyme*
- Les services AnaEE-France mobilisés*
- Le projet de recherche (utile pour associer des projets d'expérimentation mobilisant plusieurs services)
- Un descriptif résumé du projet : objectifs, conception de l'expérience ou des expériences réalisées, résultats attendus
- Un résumé grand public (optionnel) *
- Le type d'écosystème
- Les variables mesurées
- Les facteurs manipulés
- Dates d'exécution sur la plateforme et d'exécution du projet
- Responsable du projet* (attention à bien respecter les clauses RGPD en ajoutant une note de clauses standard et un lien à cliquer par l'utilisateur)
- Liens vers les publications (à actualiser au fil de l'eau)

* champs diffusés dès l'acceptation du projet sous réserve d'accord du responsable du projet.

Périmètre d'un jeu de données acquis sur une plateforme expérimentale

Le cadre général

Un jeu de données produit par une expérience sur une plateforme expérimentale AnaEE-France est composé de données produites par les plateformes et de données produites par les utilisateurs (DU). Dans les données produites par les plateformes, on distingue les données de base (DB) caractérisant l'environnement du site (météo, sol …) et les données spécifiques à l'expérience collectées par la plateforme (DP) pour le compte de l'utilisateur (mesure de gaz dans une cellule, température d'un bassin …). **Le jeu de données expérimental est alors constitué de l'union des DB (pertinentes pour l'expérience), DP et DU.**

Dans leur cycle de vie, les données font l'objet d'un certain nombre de transformations qu'il convient de bien préciser dans le jeu de données. Ainsi de manière générique on peut préciser plusieurs niveaux dans le traitement de la donnée comme dans d'autres infrastructures du domaine de l'environnement :

- L0 : données brutes (tensions délivrées par un capteur, compte numérique, proxy). Cette étape peut contenir des filtrages standards.
- L1 : données mise à l'échelle dans l'unité des grandeurs cibles (exemple : m/s pour un vent) en utilisant les procédures constructeurs. L'objet de ce niveau est de produire rapidement des valeurs ayant un sens physique, chimique ou biologique. Cette étape peut contenir des filtrages standards (type gamme de mesure).
- L2 : données calibrées et filtrées. Les données de niveaux L2 intègrent l'ensemble des procédures de filtrage et d'étalonnage mis en œuvre sur les données délivrées (filtrage automatique et manuel, étalonnage externe, correction des facteurs externes influençant la mesure). Il n'y a pas à ce stade de reconstitution de mesure. La donnée est accompagnée d'un indice de qualité qui peut être différent d'une donnée à l'autre (par exemple, covariables pas toujours disponibles, options de traitement variables, confiance en l'étalonnage, bruit sur la mesure…).
- L3 : produits élaborés pour faire intervenir un ensemble de capteurs et de modèles pour arriver à un résultat consolidé. Les traitements peuvent être, par exemple :
    - du gapfilling : on utilise une technique permettant de combler les données manquantes;
    - des agrégations de capteurs: on produit une grandeur qui tient compte de répétitions, comme pour l'évaluation d'un stock hydrique à partir de plusieurs mesures sur des capteurs indépendants,
    - assimilation de données dans un modèle : les données sont produites par un modèle dans lequel des données ont été assimilées;
    - ou reconstitution d'une donnée non directement mesurée: par exemple, les grandeurs sont issues d'un bilan.

Un jeu de données peut contenir plusieurs niveaux de traitement mentionnés ci-dessus.

Les métadonnées nécessaires à la compréhension des données et à leur exploitation doivent être associées aux données. Deux types de métadonnées sont à considérer :

- des métadonnées sur le jeu de données pour « le porter à connaissance ». Ces métadonnées ont pour vocation à alimenter les catalogues de données. Elles respecteront donc les standards internationaux afin de faciliter leur moissonnage par les catalogues nationaux et internationaux. Ces métadonnées de découverte devront renseigner la nature du jeu de données (contexte expérimental, écosystèmes étudiés, contenu du jeu de données, emprise spatiale et temporelle, contacts et droits sur l'utilisation). Ces métadonnées peuvent en partie être produites à partir des métadonnées sur les projets (voir section précédente)
- des métadonnées permettant de décrire précisément les données (métadonnées métier) en donnant pour chaque variable ou type d'observation des informations sur :
    - l'objet mesuré (quoi, où, étendue spatiale, échantillonnage, contexte expérimental (traitement)),
    - la grandeur considérée (grandeur, unité)
    - les caractéristiques temporelles (représentativité temporelle de l'information)
    - la provenance (qui ?, méthodes de mesure et de traitement).
    - et le propriétaire

Le périmètre du jeu de donnée sera précisé dans le plan de gestion de données, document définissant l'ensemble des modalités et moyens mis en œuvre pour la gestion (depuis la collecte jusqu'à la valorisation ou la destruction) des données dans le cadre du projet et initié lors de la conception de l'expérience[1].

Les cas particuliers

Expérimentations à long terme

Dans les expérimentations à long terme, la collecte des données de bases (DB) par la plateforme est inhérente à sa mission. Les données produites par les utilisateurs constituent des données complémentaires qui ont vocation à être intégrées au corpus des données gérées par la plateforme, réutilisées par la plateforme et mises à disposition.

Plateformes analytiques et instruments partagés

La mission des plateformes analytiques et des instruments partagés est de fournir des mesures qui ont vocation à s'intégrer dans les jeux de données des expériences auxquelles ces données sont rattachées, expériences pouvant être réalisées ou non sur des plateformes expérimentales AnaEE-France. Nous considérons donc que les jeux de données doivent alors suivre les règles proposées par les financeurs et celles des dispositifs expérimentaux associés. Par contre, la capitalisation sur les analyses et les métadonnées associées permettant de caractériser les conditions de l'expérience est essentielle pour l'amélioration du service analytique/instrumental produit (par exemple, pour produire un référentiel d'interprétation). L'accès aux données et métadonnées des utilisateurs par les responsables des plateformes de même que leur utilisation doit donc être garantis pour améliorer les mesures et leur interprétation.

La loi et les recommandations aux infrastructures

En France, l'ouverture par principe des données publiques disponibles au format électronique est posée par la loi pour une république numérique 2016-1321 du 7 octobre 2016. Cela concerne en particulier les bases de données des organismes publics de recherche et les données « dont la

[1] Pour l'instant la mise en place d'un plan de gestion de données ne fait pas encore partie des procédures standard de mise en place des projets. Cela devrait l'être prochainement.

publication présente un intérêt économique, social, sanitaire ou environnemental. Les données produites dans le cadre de AnaEE-France sont donc pleinement concernées. De même, la loi prévoit un principe de libre réutilisation de ces données.

En particulier, « dès lors que les données issues d'une activité de recherche financée au moins pour moitié par des dotations de l'Etat, des collectivités territoriales, des établissements publics, des subventions d'agences de financement nationales ou par des fonds de l'Union européenne ne sont pas protégées par un droit spécifique ou une réglementation particulière et qu'elles ont été rendues publiques par le chercheur, l'établissement ou l'organisme de recherche, leur réutilisation est libre. » Enfin, toute diffusion des données doit « se fait dans un standard ouvert, aisément réutilisable et exploitable par un système de traitement automatisé ».

L'ouverture par principe des données publiques de recherche et donc leur libre réutilisation ne pourront être limités qu'en cas d'exceptions encadrées par la loi et notamment :

- le secret en matière commerciale et industrielle et le secret professionnel ;
- le secret de la défense nationale et des impératifs relatifs à la sécurité de l'Etat ou de l'établissement (ex. sécurité des systèmes d'information des administrations), PPST ;
- les droits de tiers ;
- les documents non encore librement communicables au regard du code du patrimoine.

Récemment la France s'est dotée d'un plan national pour la science ouverte[2] qui réaffirme les principes d'ouverture des résultats de la recherche. Les données de recherche produites par la recherche publique française doivent progressivement être structurées en conformité avec les principes FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable). La notion d'embargo n'est pas présente dans la loi qui prévoit une ouverture par défaut et sans délai. En revanche, des délais d'ouverture des données peuvent se justifier au regard de la qualité et de la fiabilité des données attendues au regard des bonnes pratiques de la communauté scientifique concernée et des besoins des utilisateurs futurs des jeux de données, ou de questions éthiques par exemple.

En parallèle, au niveau français, les organismes de recherche ont édité des chartes sur les Infrastructures de Recherche (IR) pour diffuser librement les données qu'elles produisent (par exemple, Charte des infrastructures INRA)[3]. Au niveau européen, les chartes sur les infrastructures[4] encouragent une politique d'ouverture des données et pour ce faire, la recherche d'accords entre utilisateurs des plateformes et les responsables des plateformes de recherche.

Politique d'ouverture des données

Principes

Les métadonnées descriptives des expérimentations sont libres d'accès dès la fin du projet. Elles sont disponibles sur le portail AnaEE-France.

Les jeux de données (DP, DU, DB) produites avec les services d'AnaEE-France et par leurs utilisateurs ont vocation à être réutilisées librement selon les principes de l'Open Data. Toute personne pourra

[2] http://cache.media.enseignementsup-recherche.gouv.fr/file/Actus/67/2/PLAN_NATIONAL_SCIENCE_OUVERTE_978672.pdf

[3] https://inra-dam-front-resources-cdn.brainsonic.com/ressources/afile/368466-19bb9-resource-charte-des-infrastructures-de-recherche-inra.pdf

[4] European Charter for Access to Research Infrastructures : Principles and Guidelines for Access and Related Services, doi:10.2777/524573

avoir un accès libre et gratuit aux données avec pour seul engagement de citer l'origine du jeu de données et la date de dernière mise à jour. Les producteurs de données (gestionnaires des services et utilisateurs des services) s'engagent à appliquer ce principe aux données qu'ils génèrent.

Un jeu de données est ouvert lorsque les données ont été traitées, vérifiées et annotées avec les métadonnées appropriées. Les utilisateurs reconnaissent que les jeux de données seront automatiquement et sans délai archivées dans les espaces de stockage d'AnaEE-France et mises à disposition du public.

Aménagement possible

L'utilisateur pourra demander au responsable de la plateforme de retarder cette ouverture pour des motifs légitimes prédéfinis par ladite plateforme, éventuellement dans son Data Management Plan (maturité vérifiabilité / traçabilité, données personnelles non encore anonymisées, raisons éthiques, …etc). Le temps nécessaire pour mener ces opérations ne pourra pas excéder 2 années à partir de la fin du projet de recherche (ci-après le « délai d'ouverture »).

Pour les expérimentations à long terme, une mise à jour régulière des jeux de données est envisagée sur une base annuelle. Le délai maximum est alors compté à partir de la date de fin de la période de mise à jour (par exemple si la période est l'année 2015, le délai est compté à partir du 31 décembre de cette année).

Pendant le délai d'ouverture, le jeu de données est uniquement accessible aux personnes associées au projet de recherche. Passé ce délai, le jeu de données contenant l'ensemble des données et métadonnées sera publié et ainsi faire l'objet d'un DOI. Le jeu est rendu accessible sur un entrepôt de données ouvert.

Les publications scientifiques ne se substituent pas au jeu de données publié. En effet, les publications ne traitent en général que d'un sous ensemble du jeu de données et ne fournissent par ailleurs pas toujours l'intégralité des informations nécessaires à la réutilisation des données.

Si un utilisateur des services est défaillant dans la diffusion des données, notamment en l'absence de restitution du jeu de données dans le délai maximum imparti, la plateforme pourra diffuser les données acquises pour le compte de l'utilisateur.

Dans le cas des plateformes analytiques et des instruments partagés (voir section ci-dessus), les responsables de ces dispositifs pourront bénéficier, quelle que soit l'origine du financement (fonds privés, fond publics), des données collectées et des métadonnées associées pour des fins d'amélioration du service rendu (enrichissement d'un référentiel par exemple).

Remarques : De telles dispositions se démarquent de celles de l'accord de consortium AnaEE-France par une ouverture accrue des données[5] pour se placer dans le cadre du plan national pour la science ouverte et de la loi pour une république numérique.

[5] 9.1.2 Mise à disposition des Données : La Prestation facturée au coût marginal entraînera pour l'Utilisateur, l'obligation de mettre à disposition les Données acquises à l'aide des moyens proposés par l'Infrastructure au terme d'un délai maximum de 5 ans après l'acquisition des Données. Le propriétaire des Données pourra demander à être associé aux travaux utilisant ses Données dans le cadre d'une collaboration sur une période de 5 à 10 ans suivant l'acquisition des Données. Au-delà, le propriétaire des Données aura l'obligation d'en prévoir le libre accès. Le bénéficiaire de ces Données aura un droit d'usage pour ses seuls besoins de recherche. L'utilisation d'un service facturé au coût complet entraînera pour l'utilisateur du service la totale liberté d'exploitation des données issues de l'utilisation.

Implémentation de la politique des données

L'implémentation de la politique s'appuie sur des documents à caractère contractuel comme des chartes d'accès aux services ou des licences sur les jeux de données, et sur les systèmes d'information assurant la mise à disposition des données.

Chartes d'accès aux services

La charte a deux rôles :

- régir les échanges de données entre la plateforme et l'utilisateur afin d'assurer la réalisation de l'expérimentation et la possibilité d'utiliser les données à des fins d'amélioration des services délivrées par la plateforme.
- rappeler à l'utilisateur les règles de diffusion des jeux de données contribuant à l'Open Data conformément au plan national pour la science ouverte et à la législation (loi Lemaire) et en accord avec la politique des données de l'infrastructure nationale.

Dans la suite, les principes généraux sont déclinés par type de plateforme et l'objet de ce chapitre est de proposer une méthodologie pour aborder les questions relatives à l'usage des données dans les chartes. La question de la diffusion des données à un tiers n'est pas l'objet des chartes, si ce n'est le rappel des principes AnaEE-France et les engagements pris par les utilisateurs pour diffuser leurs données au terme de l'expérimentation.

Nature des jeux de données et règles d'utilisation

Dans le tableau ci-dessous, la nature des jeux de données ainsi que les principes de leur utilisation sont donnés par catégorie de plateforme de l'infrastructure AnaEE-France (en vert des exemples de rédactions sont donnés en annexe 1)

Catégorie de service	Nature des jeux de données	Règles sur la fourniture et l'utilisation des métadonnées et données
Plateforme d'expérimentation à façon (Ecotrons, dispositifs semi-contrôlés et plateformes d'écologie expérimental des Nouragues et du Lautaret)	Une expérience ➔ Un jeu de données composé de 3 sous-entités : - <u>données de base</u> (DB) de la plateforme caractérisant par exemple l'environnement (météo, sol …) - <u>Données expérimentales collectées par la plateforme (DP)</u> pour le compte de l'utilisateur (mesure de gaz dans une cellule, température d'un bassin … - <u>Données expérimentales additionnelles produites et collectées par l'utilisateur (DU)</u>	1- L'utilisateur s'engage à fournir à la plateforme les métadonnées descriptives de l'expérience (écosystème, manipulation réalisée, observations réalisées) pour alimenter la base de données des projets hébergés par la plateforme. 2- Dans le respect du cadre juridique relatif au partage des données, l'utilisateur responsable scientifique d'une expérience s'engage à diffuser le jeu de données librement et sans contrepartie financière. 3- Les DU pourront être gérées par le SI de la plateforme lorsque cela sera techniquement possible. Les utilisateurs bénéficieront ainsi des services de publication des données offerts par la plateforme*. 4- La plateforme garde toute liberté pour l'usage de ses données de base (DB) et peut utiliser librement les données expérimentales

		qu'elle a collectées pour le compte de l'utilisateur (DP) pour les besoins d'amélioration du service. 5- Si l'utilisateur est défaillant dans la publication du jeu de données et sa diffusion, la plateforme pourra diffuser les données qu'elle aura acquises (DP) après le délai éventuel d'ouverture fixé avec l'utilisateur, ainsi que toutes les données mises à disposition via la publication d'articles scientifiques
Plateformes d'expérimentation à long terme (dispositifs in natura)	Données de bases produites par la plateforme**(DB) sur le dispositif expérimental à long terme Données additionnelles produites par l'utilisateur sur le dispositif expérimental (DU)	1- La plateforme s'engage à fournir à l'infrastructure AnaEE-France les métadonnées descriptives du dispositif expérimental de long terme (écosystèmes, manipulations réalisées, observations réalisées) pour alimenter en métadonnées la base AnaEE- France. [exemple ACBB article 2] 2- L'utilisateur s'engage à compléter les métadonnées de l'expérimentation à long terme avec les informations relatives aux DU. [Exemple ACBB : 6.1.4.3] 3- Les données DB (historiques et données en cours) sont mises à disposition des utilisateurs à partir moment où elles sont exploitables***. [Exemple ACBB : 6.1.4.2] 4- Les utilisateurs mettent leurs données (DU) à disposition de la plateforme pour leur intégration dans son SI. [Exemple ACBB : 6.1.4.3] 5- Dans le respect du cadre juridique relatif au partage des données, l'utilisateur responsable scientifique de l'acquisition des DU s'engage à diffuser le jeu de données librement et sans contrepartie financière dès leur achèvement ou au plus tard après le délai d'ouverture fixé avec la plateforme. Les DU seront diffusées selon les mêmes supports que les DB. [Exemple ACBB : 6.1.4.3]
Plateformes analytiques et instruments partagés	Un jeu de données composé de deux sous-entités : Données produites par les moyens analytiques**** (DP) Métadonnées fournies par l'utilisateur (MDU) pour l'interprétation des mesures	1- L'utilisateur s'engage à fournir à la plateforme les métadonnées descriptives de l'expérience (écosystème, manipulation réalisée, observations réalisées) pour alimenter la base de données des projets hébergés par la plateforme. 2- La plateforme dispose de l'usage des DP et des MDU pour les besoins d'amélioration du service (amélioration du référentiel) [exemple

		Biochemenv 6.1.5.1 : Contribution au référentiel du service] 3- Si les données DP sont acquises dans le cadre d'une expérimentation sur les plateformes AnaEE France, les utilisateurs s'engagent à intégrer les DP dans le jeu de données expérimental et les diffuser selon les principes de la ou les plateformes AnaEE-FR concernées. [exemple Biochemenv 6.1.5.2 Condition d'accès aux données produites par la plateforme] 4- La plateforme peut diffuser le référentiel selon les règles fixées avec l'utilisateur. [non traité dans l'exemple]

* Pour la publication des données produites par les expérimentations financées majoritairement sur fonds publics, il semblerait naturel que cela soit sous la coresponsabilité de l'utilisateur du service et de la plateforme. Il faut sans doute ici prévoir un article précisant que l'ouverture des données (conformément au cadre législatif) fait l'objet d'un engagement de l'utilisateur du service et qu'en cas de défaillance, la plateforme ou l'infrastructure est en droit/devoir de les publier. L'utilisateur pourrait aussi, d'emblée, transférer à la plateforme (si elle dispose de ce service) la responsabilité de publier les données.
** Notamment toutes les données du dispositif expérimental de base des expérimentations à long terme des services SOERE.
*** lorsque la donnée brute a subi les traitements nécessaires pour être compréhensible, interprétable et exploitable
**** Données produites avec les ressources humaines de la plateforme ou par celles de l'utilisateur

Licences

Les jeux de données produits pourront faire l'objet d'une licence. La licence à utiliser pour les données produites par les services de AnaEE-France devrait être la 'Licence Ouverte' ((licence LO 2.0, ODBL équivalente à la licence CC-BY). L'ensemble des activités d'AnaEE-Fr et en particulier la production des données s'inscrivant d'emblée dans un cadre international, l'utilisation de la licence CC-BY 4.0 serait la licence recommandée.

Les Systèmes d'Information (SI)

Les SI associés aux plateformes d'AnaEE-France jouent un rôle important pour la diffusion des données, mais la publication des jeux de données conduit à les mettre à disposition via les entrepôts de données. Deux voies de diffusion sont à considérer.

Dans AnaEE-France, deux familles de SI sont en cours de développement :

- Le SI Ecoinfo a vocation à rassembler les DB des expérimentations à long terme ainsi que les données complémentaires (DU) des utilisateurs. Le SI offre ainsi plusieurs services :
  - Un environnement de stockage propre et sécurisé ;
  - Une interface de requêtes pour extraire des jeux de données à façon ;
  - Une gestion des droits et d'attribution des licences ;
- Le SI ISIA-EDEM est en cours de développement et destiné aux expérimentations à façon. Il aura dans un premier temps les mêmes fonctions que le SI Ecoinfo pour les données de DB et DP. Un service aux utilisateurs pourra être fourni pour gérer également les DU et ainsi obtenir un environnement unique pour gérer l'ensemble du jeu de données d'une expérience et faciliter l'accès aux données et leur publication à la fin du processus.

Le SI global d'AnaEE-France valorise ces SI 'locaux' en développant leur l'interopérabilité au travers de l'annotation sémantique des ressources. Cette annotation utilise une ontologie partagée basée sur

OBOE. Des pipelines informatiques sont développés pour automatiser les traitements sémantiques. Un premier est consacré à l'annotation et à la production des données et métadonnées. Les métadonnées décrivent les ressources (ici données) gérées dans les SI locaux et alimentent le portail AnaEE-France d'accès à ces ressources. Un deuxième pipeline est consacré à l'exploitation des données sémantiques à travers la génération i) de métadonnées ISO et GeoDCAT normalisées et ii) de fichiers de données (NetCDF) à partir de périmètres sélectionnés (sites expérimentaux, années, facteurs expérimentaux, variables mesurées...). Ces pipelines sont développés dans le contexte d'ENVRIplus et contribuent à son portefeuille de services. La généricité des outils doit garantir leur réutilisation dans différents contextes d'ontologies et de bases de données.

L'utilisation des SI locaux pour diffuser les données est un environnement plutôt approprié. Il souffre néanmoins d'être en perpétuelle évolution et il peut être difficile de raccorder une extraction à un DOI qui fige l'état d'une base à un état donné. Il faut donc établir une stratégie qui tienne compte des différentes voies de diffusion avec néanmoins un principe partagé qui est le référencement avec un DOI de toutes les données récupérées par un utilisateur, quelle que soit la voie de diffusion. On peut ainsi imaginer plusieurs scénarios (liste non exhaustive) :

- Les jeux de données publiés et déposés dans les entrepôts constituent la seule voie de diffusion ouverte des données. Les SI des plateformes demeurent un outil interne permettant de faciliter le travail des utilisateurs et les gestionnaires des plateformes. Les jeux de données déposés dans les entrepôts sont produits soit directement à partir des SI soit grâce au pipeline d'exploitation des données sémantiques qui offre un service de génération du jeu de données (extrait depuis le SI de la plateforme), d'attribution d'un DOI et de dépôt sur un entrepôt.
- Le SI dans son ensemble est figé et publié lorsque le jeu de données est achevé. Toutes extractions du SI doivent alors faire référence au DOI associé. On peut imaginer un versionnage du SI (en particulier pour les expérimentations à long terme) incluant les mises à jour. Cela suppose d'avoir en permanence deux versions du SI, une de travail dans lequel se font les mises à jour et une version ouverte et figée qui est rattachée au DOI.
- Les deux systèmes précédents co-existent avec l'inconvénient qu'une même donnée peut être rattachée à deux DOI.

Annexe 2 : métadonnées associées aux indicateurs biologiques mesurés par la plateforme

Métadonnées obligatoires

-numéro d'identification

-date de prélèvement

-nom du projet

-site échantillonné (nom, coordonnées géographiques)

-identification de la parcelle (nom, itinéraire technique)

-type d'échantillon (sol, invertébré…)

-horizon échantillonné (profondeur min et max)

Métadonnées optionnelles

-texture du sol, pH, taux de matière organique

-conditions de prélèvement (humidité du sol, état du couvert)

-espèce (pour les organismes)

Annexe 3 : Acceptation de la charte et des modalités d'accès à la plateforme Biochem-Env

Je soussigné (nom, prénom, organisme) :

Projet/prestation (dénomination/acronyme) :

Déclare avoir lu la charte d'accès à la plateforme Biochem-Env et s'engage à en respecter les termes.

Fait à : **Le :**