

Politique de libre accès aux publications et aux données

Les données de phénotypage sont coûteuses et ne peuvent pas être directement reproduites car les combinaisons de conditions environnementales et de pratiques expérimentales ne peuvent jamais être reproduites à l'identique, sauf pour quelques expériences dans des chambres de culture de haute précision. Les recherches de Phenome-EMPHASIS ont donc vocation à être réanalysées par la communauté scientifique, après avoir été analysées par les équipes qui ont produit les données. Phenome-EMPHASIS met donc tout en œuvre pour que les données produites soient Faciles à trouver, Accessibles, Interopérables et Réutilisables (FAIR, "Findable, Accessible, Interoperable, Reusable" en anglais). Phenome-EMPHASIS a mis pour cela au point des outils et des règles spécifiques présentées dans le plan de gestion des données, lui-même dépendant de l'infrastructure européenne EMPHASIS. La charte est compatible avec les principes énoncés dans l'infrastructure européenne EOSC et dans le document "charte pour le libre accès aux publications et aux données"

<https://datapartage.inrae.fr/Documents-de-referance/Charte-Open-Access-de-l-INRA>

Les principes énoncés ici sont détaillés dans le plan de gestion des données de Phenome-EMPHASIS
<https://www.phenome-emphasis.fr/Acces/Gestion-des-donnees>

1. Principes généraux

La présente politique de gestion des données porte sur la propriété et la publication des données expérimentales et des métadonnées associées collectées dans le cadre de Phenome-EMPHASIS, ainsi que sur l'accès à ces données et métadonnées.

L'acceptation de cette politique est une condition nécessaire à l'appartenance d'une installation à l'infrastructure nationale Phenome-EMPHASIS, et à l'accès d'un utilisateur à une des installations de l'infrastructure.

2. Définitions

Les termes infrastructure nationale, infrastructure locale, installation, facility, expérience, projet sont définis sur la Figure 1.

Le terme **données expérimentales** désigne les données collectées à partir d'une expérience réalisée sur une installation de l'infrastructure Phenome-EMPHASIS. Cette définition comprend les données collectées automatiquement par les capteurs, ou manuellement par le personnel de la plateforme ou les utilisateurs de la plateforme.

Les **Métadonnées** sont les informations portant sur les données collectées nécessaires à la gestion et l'analyse des données expérimentales, par exemple sur le contexte de l'expérience, les pratiques expérimentales, la position des plantes à chaque instant ou la calibration des capteurs.

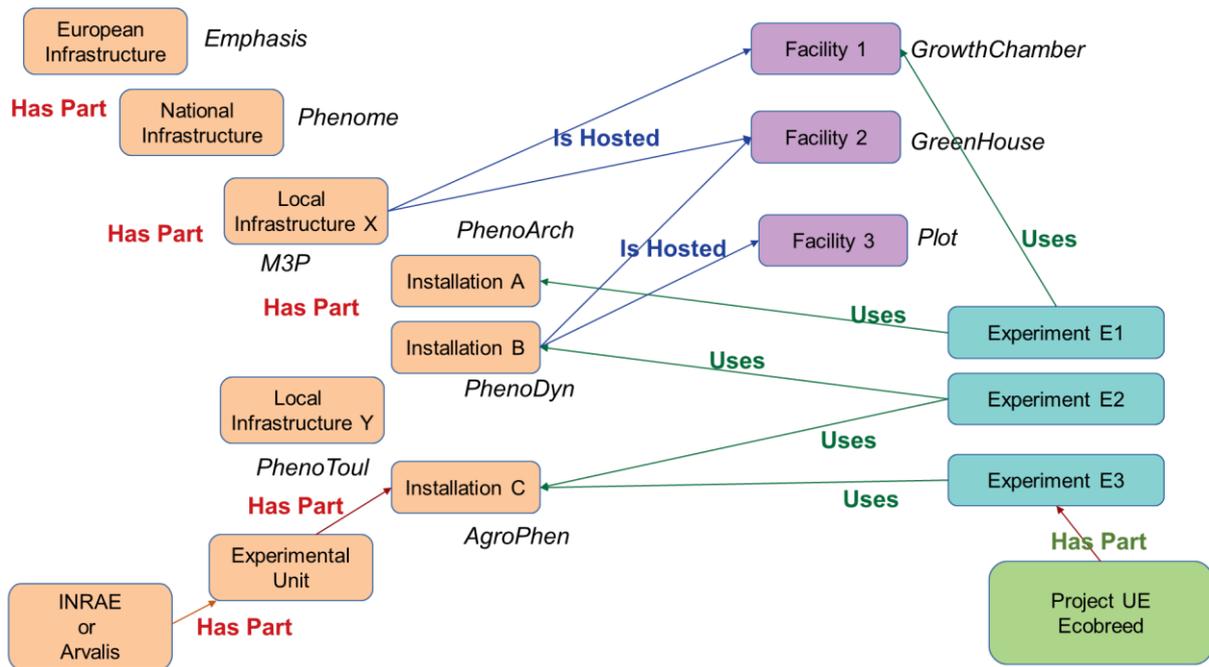


Figure 1. Organisation des infrastructures et expériences. Emphasis est l'infrastructure européenne, qui implique des infrastructures nationales elles-mêmes formées d'infrastructures locales. Elles impliquent aussi des "facilities" qui peuvent héberger une installation ou être utilisées dans des expériences en plus des installations (ex. Chambre de vernalisation). Une installation peut être une partie d'une unité expérimentale (sans lien 'has part' entre UE et Phenome-Emphasis), elle-même partie d'un organisme. Une expérience fait partie d'un projet et utilise une installation et des 'facilities'. Tous ces liens sont documentés via des URIs.

Un DOI (**Digital Object Identifier**) est un identifiant persistant pour un objet, généré par un institut enregistré auprès de datacite comme l'INRAE. Toute donnée référencée par un DOI doit être disponible pendant 10 ans minimum de manière raisonnablement stable. Les DOI sont typiquement utilisables pour un jeu de données issues d'une expérience, au moment de sa publication, ou pour identifier le matériel végétal. Un URI (**Uniform Resource Identifier**) est un identifiant persistant pour une expérience et pour tout objet utilisé dans une expérience (plante, pot, vecteur, capteur).

Les **jeux de données** sont les combinaisons d'informations collectées pendant une expérience. Ils sont de deux types principaux, gérés différemment par l'infrastructure et par les utilisateurs.

- **Les données phénotypiques correspondant à une expérience individuelle** recouvrent les images, sorties de capteurs environnementaux, sorties de capteurs sur plantes, sorties d'analyseurs d'échantillons et observations collectées pendant l'expérience, conjointement avec les métadonnées qui les accompagnent. La quasi-totalité de ces données est fonction du temps (par exemple température relevée toute les heures ou biomasse chaque jour). Elles sont organisées dans un système d'information, par défaut PHIS (Neveu et al 2019) ou système équivalent, qui relie tous ces éléments à leurs métadonnées et les rendent FAIR. L'organisation et le maintien de ces jeux de données sont sous la responsabilité de Phenome-EMPHASIS.
- **Les données synthétiques analysées pour une ou plusieurs expériences** (par exemple les moyennes génotypiques corrigées intra ou inter expériences ou des indicateurs environnementaux comme la température moyenne journalière pendant la floraison), sont organisées en suivant les règles mises en place par le groupe de travail Emphasis – Elixir

MIAPPE (<https://www.miappe.org/>) dans des entrepôts compatibles avec l'infrastructure ELIXIR (par exemple GNPIS ou Data-INRAE).

Le terme **accès ouvert** signifie non protégé par un copyright ou un brevet, et susceptible de réutilisation libre, par exemple sous la licence CC-BY <https://creativecommons.org/licenses/by/4.0/>.

Le terme **Période d'embargo** signifie la période, à partir de la fin de l'expérience, pendant lequel l'accès aux données collectées et aux métadonnées est limitée aux bénéficiaires de l'accès. Pour les projets académiques, elle est définie dans les plans de gestion des données des projets qui ont sollicité l'expérience et ne peut dépasser cinq (5) ans sauf explicitement mentionné dans la convention d'accueil. Pour les projets industriels, elle est définie lors de la rédaction de la convention d'accueil et ne peut dépasser dix (10) ans.

3. Propriété

3.1. Les données expérimentales et les métadonnées obtenues par des recherches menées dans les installations de Phenome-EMPHASIS seront automatiquement en accès ouvert après la période d'embargo.

3.2. Les jeux de données phénotypiques pour les expériences particulières et les jeux de données synthétiques appartiennent aux chercheurs qui ont bénéficié de l'accès à l'installation (et à travers eux aux organismes auxquels ils appartiennent). Il s'agit des images, des mesures et des observations réalisées pendant les expériences, ainsi que les résultats des analyses qui ont suivi l'expérience.

Ces jeux de données sont gérés par les Plans de Gestion des Données des projets (par exemple ANR ou UE) qui ont sollicité l'accès. Les règles spécifiques de Phenome-EMPHASIS sont :

- Les jeux de données des expériences individuelles sont organisés suivant les principes énoncés par PHIS (identification, traçabilité, origine, Neveu et al 2019) et chaque fois que possible organisés dans le système d'information PHIS. Le dépôt de données sur les sites des journaux est fortement déconseillé. De même, le dépôt sur des systèmes gratuits comme CyVerse, dont la permanence n'est pas garantie et ne respectant pas les principes PHIS et MIAPPE, est incompatible avec Phenome-EMPHASIS.
- Les jeux de données publiés doivent avoir une licence ouverte de réutilisation sous réserves des clauses de l'accord de consortium et du plan de gestion des données du projet utilisateur.

3.3 Les jeux de données issus de projets technologiques seront publiés chaque fois que possible. En particulier, les PME ne sont pas obligées de diffuser leurs résultats méthodologiques à un large public, mais sont encouragées à le faire après brevet éventuel.

3.4 Quatre catégories de données restent la propriété des installations dans lesquelles les expériences ont eu lieu, et mises à disposition des bénéficiaires d'accès : (i) les données environnementales collectées pendant les expériences, pour que les responsables d'installations puissent réaliser des méta analyses de ces données. (ii) Les procédures et résultats des étalonnages des capteurs et des caméras. (iii) Les chaînes d'analyse et les procédures utilisées pour extraire les données phénotypiques à partir de capteurs et d'images. (iv) Les innovations et logiciels développés pour optimiser les plans d'expérience dans l'installation, ainsi que les outils statistiques pour analyser les données en particulier détections d'outliers et analyse spatiale.

4. Rôles et responsabilités

- Sauf indication contraire, Phenome-EMPHASIS agira comme dépositaire des données et métadonnées des expériences individuelles, via PHIS ou système équivalent.

- L'équipe qui accueille l'accès s'assurera que les métadonnées sont aussi complètes que possible, pour faciliter la recherche, la récupération et interprétation des données à long terme.
- Les projets qui sollicitent les expériences doivent établir un Plan de Gestion des Données.
- Les bénéficiaires de l'expériences définissent et transmettent à l'infrastructure locale les autorisations d'accès, ainsi que les éventuelles codifications de noms des géotypes.

5. Informations relatives aux publications

- Les publications relatives aux données des expériences réalisées dans une installation devront citer l'infrastructure Phenome-EMPHASIS via son DOI ou (transitoirement) le numéro de projet ANR-INBS, l'URI de l'expérience et des données, ainsi que le DOI de l'installation.
- Phenome-EMPHASIS encourage les bénéficiaires d'accès à placer les jeux de données synthétiques dans des entrepôt institutionnel et/ou dans des entrepôts compatibles avec l'infrastructure ELIXIR, et à publier les résultats dans des revues "open access".