



Implémentation d'un PGD entité/structure : exemple de la plateforme ICS (Institut Clinique de la souris)

Hamid Meziane, PhD, Phenomin-ICS
meziane@igbmc.fr

Printemps des plateformes, Strasbourg
28/03/2024



Contexte

- La science ouverte est la diffusion sans entrave des résultats, des méthodes et des produits de la recherche scientifique.
 - ❖ Elle s'appuie sur l'opportunité que représente la mutation numérique pour développer l'accès ouvert aux publications et – autant que possible – aux données, aux codes sources et aux méthodes de la recherche.
- 2e axe du 2e plan national pour la science ouverte 2021-2024
 - ❖ Structurer, partager, et ouvrir les données de la recherche
- Mesures
 - ❖ **Mettre en œuvre l'obligation de diffusion des données de recherche** financées sur fonds publics
 - ❖ **Créer Recherche Data Gouv**, la plateforme nationale fédérée des données de la recherche
 - ❖ **Promouvoir l'adoption d'une politique de données** sur l'ensemble du cycle des données de la recherche, pour les rendre faciles à trouver, accessibles, interopérables et réutilisables ([FAIR](#))

Contexte

- 4e axe du 2e plan national pour la science ouverte 2021-2024

Transformer les pratiques pour faire de la science ouverte le principe par défaut

- S'appuyer sur les infrastructures inscrites dans la feuille de route nationale pour transformer les pratiques et généraliser la science ouverte, en les invitant à:

- ❖ Formaliser leur politique de science ouverte en rendant public un document stratégique, et à **mettre en oeuvre de manière effective** les principes FAIR et **les plans de gestion de données**, par et pour leurs utilisateurs
- ❖ **Inscrire dans leurs conditions d'accès** les principes d'ouverture des publications, et d'ouverture par défaut des données et des codes sources

Le partage de données de la recherche e.g. à travers un entrepôt, permet d'augmenter leur traçabilité en répondant de façon implicite au concept FAIR, un des piliers de la Science Ouverte.



S'APPUYER SUR LES INFRASTRUCTURES DE RECHERCHE INSCRITES DANS LA FEUILLE DE ROUTE NATIONALE POUR TRANSFORMER LES PRATIQUES ET GÉNÉRALISER LA SCIENCE OUVERTE

- Inviter les infrastructures à inscrire dans leurs conditions d'accès les principes d'ouverture des publications et d'ouverture par défaut des données et des codes sources.
- Inviter les infrastructures à formaliser leur politique de science ouverte en rendant public un document stratégique, et à mettre en œuvre de manière effective les principes FAIR et les plans de gestion de données, par et pour leurs utilisateurs.
- Doter chaque infrastructure d'un identifiant unique (funder ID) que les chercheurs seront appelés à mentionner dans les publications, codes et données produits grâce à cette infrastructure.
- Encourager les tutelles des infrastructures dans une politique de recrutement de métiers chargés du traitement, de la qualité, de la description et de la préservation des données.



Vue aérienne de l'Observatoire du Pic du Midi

- Mobiliser les infrastructures pour constituer et porter les grands jeux de référence ainsi que les référentiels de pratiques disciplinaires et thématiques de préservation, de description et de référencement des données de recherche.
- Mettre en œuvre les projets lauréats de l'appel à manifestation d'intérêt Equipements structurants pour la recherche du PIA qui renforcent le développement d'infrastructures, plateformes et services de données thématiques.



PGD ?

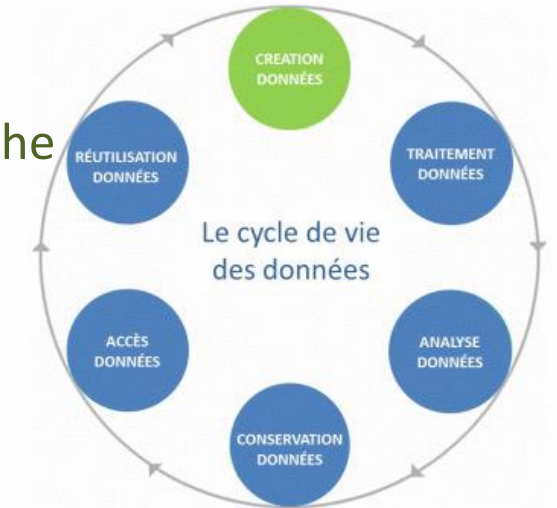
- Pourquoi un PGD ? C'est quoi un PGD ?
- Réutiliser les données? Où est le problème?

https://youtu.be/66oNv_DJuPc





- Un PGD sert à documenter le cycle de vie des données d'un projet de recherche
- Il contextualise les données
 - ❖ En rassemblant les informations techniques, documentaires, administratives et juridiques
 - ✓ Permettant un **partage et un archivage** simplifié
- Le PGD est demandé par les agences de financement de la recherche (ANR, Horizon Europe)
- Il augmente la **reproductibilité et l'intégrité** de la recherche
- Il peut être utilisé comme **outil d'accompagnement à la gestion de projet** (planification)





➤ Horizon Europe et Euratom

- ❖ Rédiger et mettre à jour régulièrement un PGD
 - ✓ 1ere version à rendre dans les 6 mois du projet
 - ✓ Un PGD complet à la fin du projet de recherche
 - ✓ Si évaluation intermédiaire, PGD mise à jour à fournir
- ❖ Les données FAIR soient déposées dans un entrepôt de données de confiance
 - ✓ Aussi vite que possible
 - ✓ Selon les modalités indiquées dans le PGD
 - ✓ En libre accès « open access » (les exceptions à l'open access sont à justifier dans le PGD)
- ❖ Utilisation de la licence CC-BY ou CC-0 ou équivalent pour les données ouvertes
- ❖ Des informations via l'entrepôt sur tout autre produit de la recherche, instruments, outils nécessaire pour réutiliser ou valider les données
- ❖ Des exigences en matière de métadonnées
- ❖ Les coûts associés à la gestion des données sont éligible au remboursement



Obligations des financeurs

➤ Agence nationale de la recherche (ANR)

❖ Un PGD

- ✓ 1ere version à rendre dans les 6 premiers mois du projet
- ✓ Un PGD complet à la fin du projet de recherche
- ✓ Pour projet >30 mois, PGD mise à jour à mi-parcours

❖ Dépôt des données dans un entrepôt (aussi ouvert que possible, aussi fermé que nécessaire)

❖ Les coûts associés à la gestion des données sont éligible au remboursement

- ✓ acquisition, collecte, stockage, personnel dédié...
- ✓ Coût de stockage des données en cas de recours à un tiers **jusqu'à 5 ans** après la date de fin scientifique du projet, si contrat conclu avant la fin du projet

Contenu d'un PGD

➤ Responsabilités données

- ❖ qui a créé les données ? Qui est responsable de leurs gestion, de leur qualité? Les coordonnées sont elles indiquées

➤ La collecte des données

- ❖ Sources, formats de fichiers, volumétrie, méthodes,...

➤ Métadonnées et la documentation

- ❖ Documentation des données ?
- ❖ Quelles sont les standards de métadonnées utilisées pour décrire les données?
- ❖ Arborescence des fichiers et convention de nommage ?

➤ Le stockage :

- ❖ pendant le projet (données chaudes et tièdes),
- ❖ Après le projet (données froides)



Contenu d'un PGD

➤ Ethique

- ❖ Si les données sont sensibles? le projet a-t-il été présenté à un comité d'éthique?

➤ Juridique

- ❖ Les données sont-elles soumises au droit d'auteur ou un autre droit pouvant restreindre leur dissémination ? Les raisons de définir un embargo ?

➤ Partage et réutilisation

- ❖ Les données seront-elles déposées dans un entrepôt de données ? Sous quelle licence ? Si non, pourquoi?

➤ Archivage

- ❖ Quelle est la stratégie prévue pour la conservation et l'accès à long terme aux données ?

➤ Ressources

- ❖ Quelle sont les moyens alloués à la gestion des données ?





Comment rédiger un PGD ?

- PGD doit être pris en compte dès le début du cycle de vie du projet.
- DMP OPIDOR (basé sur la feuille de route du « Digital curation centre et le centre de curation de l'Université de Californie »
 - ❖ Personnalisé par l'INIST pour les chercheurs français
- DSW (Data Stewardship Wisard)
 - ❖ Né d'une collaboration entre ELIXIR-CZ (CTU à Prague) et ELIXIR-NL (DTL)
 - ❖ Recommandé par Horizon Europe ainsi que par plusieurs agences de financement nationales européennes
- PGD projet vs PGD Structure
 - ❖ Un projet de « PGD structure », porté par et réalisé conjointement avec les INBS, FBI, IFB, et « Centre National De Ressources Biologiques Marines »(EMBRC-France); GT piloté/animé par **Paulette Lieby et Frédéric Brault**



- PGD structure CELPHEDIA-ICS qui a été élaboré dans DSW
- C'est la formalisation du processus de gestion de nos données décrivant les moyens permettant de les rendre visibles, accessibles, et réutilisables quel que soit le demandeur.
 - ❖ Basé sur une architecture de questions, destinée à moissonner les informations relatives aux données de recherche, et permet que les champs idoines du PGD projet soient renseignés.
 - ❖ Décrit les éléments du cycle de vie de os données, c'est-à-dire, les processus de création, d'identification, de documentation, de partage et d'archivage.
 - ❖ Il s'inscrit dans la démarche de la science ouverte qui répond aux principes FAIR " Facile à découvrir, Accessible, Interopérable, Réutilisable".
 - ✓ Ce PGD structure a pour vocation de servir ensuite de modèle à tout PGD de projet porté par l'infrastructure ou le centre.



DSW@IFB

dsw.france-bioinformatique.fr/wizard/projects/4c1c3dd8-466c-4bf7-a1b7-fbce3ef7d8b0

Plan de gestion de données PHENOMIN-ICS

Questionnaire Metrics Preview Documents Settings

View Import replies Warnings 3 TODOs Comments Version history

Current Phase
Before Submitting the Proposal

Chapters

- I. Introduction ✓
- Quel est le domaine disciplinaire ...
- II. Informations générales ✓
- III. Données de la recherche ✓

I. Introduction

Ce "PGD structure" a pour objectif d'accompagner l'Institut Clinique de la Souris (ICS) et ses utilisateurs dans l'élaboration de leur PGD projet. Ce document décrit les éléments du cycle de vie de la donnée, c'est-à-dire, les processus de création, d'identification, de documentation, de partage et d'archivage. Il s'inscrit dans la démarche de la science ouverte qui répond aux principes FAIR " Facile à découvrir, Accessible, Interopérable, Réutilisable"

✓ I.1 Quel est le domaine disciplinaire concerné ?

PGD_par_defaut

Créez autant de blocs de réponses que nécessaire si plusieurs domaines sont concernés.

Desirable: *Before Finishing the Project*

- a. Multidisciplinaire (Génération et caractérisation d'animaux modèles rongeurs)
- b. Exploration fonctionnelle
- c. Ingénierie génétique

Clear answer

Answered 4 months ago by Hamid MEZIANE.

L'Institut Clinique de la Souris - ICS (Mouse Clinical Institute - MCI) est une infrastructure française de recherche d'excellence en recherche translationnelle et en génomique fonctionnelle. Elle propose un ensemble complet de services spécialisés aux utilisateurs académiques et industriels. Les activités de l'ICS incluent la génération/production de modèles rat et souris génétiquement modifiés ou environnementaux (chirurgicaux, pharmacologiques, diètes variées, ...), l'exploration fonctionnelle (phénotypage) dans différents domaines biologiques et les pathologies associées (Le métabolisme et la biochimie, le système nerveux, le système cardiovasculaire, le système respiratoire, le système immunitaire, ...) et l'archivage et la distribution d'animaux et d'échantillons biologiques. Pour mener à bien ses activités, l'ICS comporte 3 principaux départements : l'ingénierie génétique, l'animalerie et le phénotypage.

Hamid MEZIANE
Data Steward

Collapse sidebar

Taper ici pour rechercher

9°C 11:04 27/03/2024



Domaine disciplinaire?: Multidisciplinaire

- L'ICS est une infrastructure de recherche d'excellence en recherche translationnelle et en génomique fonctionnelle.
- Elle propose un ensemble complet de services spécialisés aux utilisateurs académiques et industriels.
- Les activités de l'ICS incluent
 - ❖ la génération/production de modèles rat et souris génétiquement modifiés ou environnementaux
 - ❖ L'exploration fonctionnelle (phénotypage) : Le métabolisme et la biochimie, le système nerveux, le système cardiovasculaire, le système respiratoire, le système immunitaire, ...)
 - ❖ et l'archivage et la distribution d'animaux et d'échantillons biologiques.
- Les données concernées dans ce PGD sont celles :
 - ❖ Relatives à la création et la validation d'un animal génétiquement modifié, des données de quantification d'ADN ou ARN par PCR digitale, des données de séquençage transcriptomique et génomique, ou des données de séquençage métagénomiques (ADN 16S).
 - ❖ Issues des explorations fonctionnelles (phénotypage) dans différents domaines biologiques (métabolisme et biochimie, système nerveux, système cardiovasculaire, système respiratoire, système immunitaire, ...)



II. Informations générales

1. Informations sur la structure

Nom de la structure

✓ Institut Clinique de la Souris/Mouse Clinical Institute

Acronyme

✓ a. Phenomin-ICS

URL

✓ <http://www.ics-mci.fr/en/>

Directeur :

✓ Yann Hérault ; herault@igbmc.fr

Type de structure|

✓ Plateforme d'acquisition



La structure (ou service) fait-elle partie d'une infrastructure ou une structuration locale/régionale/nationale/internationale ?

Infrastructure nationale

- ✓ CELPHEDIA (Création, ELevage, PHEnotypage, DIstribution et Archivage d'organismes modèle)
- ✓ PHENOMIN (Infrastructure Française pour la génomique fonctionnelle chez la souris)

Infrastructure internationale

- ✓ INFRAFRONTIER

Tutelles de la structure

- ✓ Université de Strasbourg
- ✓ CNRS
- ✓ INSERM

S'il y a lieu, indiquer le ou les identifiant (s) de la structure ou service ?

- ✓ ROR : <https://ror.org/03cjqqq10>
- ✓ ISNI: ISNI 0000 0004 0404 8159
- ✓ PRID: RRID:SCR_011021
- ✓ GRID: grid.452426.3

Label(s) et certification(s)

- ✓ GIS IBISA
- ✓ ISO 9001
- ✓ NFX 50-900



2. Personnes et/ou entités assumant des rôles et responsabilités qui sont pertinents à la gestion des données

- ✓ Services de l'ICS, responsables et leurs équipes techniques

Email

- ✓ ics@igbmc.fr

Pour quel rôle ou responsabilité ?

- ✓ Gestion des accès aux données pendant le projet
- ✓ Sauvegarde des données pendant le projet
- ✓ Archivage à long terme des données

3. Information sur le PGD

Le PGD a-t-il ou aura-t-il un identifiant ?

- ✓ Non

Quelles sont la/les personne(s) impliquée(s) dans la rédaction du PGD ?

- ✓ Responsables de services ; Responsable gestion des données

Une licence est-elle ou sera-t-elle attribuée au PGD ?

- ✓ Oui, CC-BY-SA



III. Données de la recherche

1. Les produits de la recherche

Nom

- ✓ Données de création et/ou de caractérisation de modèles animaux rongeurs

Description succincte

- ✓ Données de création et la validation d'un animal génétiquement modifié, des données de quantification d'ADN ou ARN par PCR digitale, des données de séquençage transcriptomique et génomique, ou des données de séquençage métagénomiques (ADN 16S).

Donnes d'explorations fonctionnelles (phénotypage) dans différents domaines biologiques et les pathologies associées (Le métabolisme et la biochimie, le système nerveux, le système cardiovasculaire, le système respiratoire, le système immunitaire, ...)

Le type du(des) produit(s) de la recherche

- ✓ Audiovisuel
- ✓ Jeux de données d'exploration fonctionnelle
- ✓ Jeux de données génomiques
- ✓ Jeux de données transcriptomiques
- ✓ Image
- ✓ Objet physique
- ✓ Son
- ✓ Texte
- ✓ Workflow



2. Description des produits de la recherche

Par qui (entité) sont générées les données ?

- ✓ Toutes les données sont générées par la structure

Génération des données

Si pertinent, quel est le niveau de traitement des données ?

- ✓ Données brutes
- ✓ Données primaires
- ✓ Données analysées

Modalités d'acquisition

- ✓ Observation, annotation manuelle.
- ✓ Acquisition automatisée

Détails répertoriés dans fichiers annexes/métadonnées.

Formats de fichiers utilisés quelque soit le type de données

- ✓ OME-TIFF; JPG; PNG; TIFF; DCAT; TXT; CSV; PDF; XLSX; DOCX; MPG; MP4; AVI; SQLITE; FASTQ; FATSAs; SAM; BAM; GenBank; autres constructeurs



➤ Métadonnées

❖ Métadonnées décrivant [l'acquisition des données \(équipements et conditions d'acquisition\)](#)

OPEN FIELD	Version	Req. Upload	Req. Analysis	Unit	Type	CENTRE SUGGESTED OPTIONS
Equipment ID IMPC_OFD_023_001	1	✓			TEXT	NI-01/557
Equipment manufacturer IMPC_OFD_024_001	1	✓	✓		TEXT	Panlab
Equipment model IMPC_OFD_025_001	1	✓	✓		TEXT	LE8822
Central zone surface area IMPC_OFD_026_001	1	✓	✓	% of total area	INT	40% of whole area
Light intensity in the centre of the arena IMPC_OFD_027_001	1	✓		Lux	FLOAT	150
Number of animals per cage IMPC_OFD_028_001	1				INT	variable
Color of arena IMPC_OFD_029_001	1				TEXT	Transparent but infrared frames
Height of the wall IMPC_OFD_030_001	1			cm	FLOAT	17
Distance from light source: if direct illuminaton IMPC_OFD_031_001	1			cm	FLOAT	not applicable
Periphery zone IMPC_OFD_032_001	1,1	✓	✓	cm	FLOAT	8
Type of analysis IMPC_OFD_033_001	1,1	✓	✓		TEXT	Wide beam break
Arena Size IMPC_OFD_034_001	1	✓			TEXT	45 X 45 cm
Experimenter ID IMPC_OFD_035_001	1,1	✓			TEXT	variable
Disinfectant IMPC_OFD_036_001	1,1				TEXT	Alcohol
Start Time IMPC_OFD_037_001	1,1				DATETIME	variable
Arena ID IMPC_OFD_038_001	1				INT	not applicable
Software Version IMPC_OFD_039_001	1	✓			TEXT	Actitrack system
Arena Material IMPC_OFD_040_001	1				TEXT	PVC
Date equipment last calibrated IMPC_OFD_041_001	1				DATETIME	before each assay



Documentation et qualité des données

Stratégies employées pour garantir la qualité des données ?

- ✓ Les protocoles d'acquisition sont décrits dans des procédures standards normalisés.
- ✓ Le personnel réalisant les procédures est formé et compétent pour leur réalisation.
- ✓ Les données brutes générées sont saisies au format informatique, avec une vérification.
- ✓ Les données brutes saisies sont ensuite vérifiées afin de s'assurer qu'elles entrent dans une plage de données physiologiques afin d'en garantir la pertinence.
- ✓ Les données sont décrites avec un vocabulaire contrôlé
- ✓ Le vocabulaire utilisé respecte les principes FAIR



La procédure est également enregistrée dans la GED (Gestion documentaire des procédures et instructions) Outil Avanteam:

❖ <https://ics.avanteam-online.com/qmdoc/ExtMain2.aspx?Nav=QDOCACTEURS>

A screenshot of the Avanteam GED interface. The browser address bar shows the URL: ics.avanteam-online.com/qmdoc/ExtMain2.aspx?Nav=QDOCACTEURS. The page header includes the 'phenomin' logo and navigation tabs: 'Accueil', 'Gestion documentaire', and 'Enregistrements'. The user 'MEZIANE Hamid' is logged in. A left sidebar contains navigation options like 'Bureau', 'Documents applicable', and 'Demande de révision'. The main content area is titled 'Par service' and features a table with columns for 'Référence', 'Titre', 'Processus', and 'Type de document'. The table lists various teams such as 'Anatomy and Histopathology team', 'Animalerie logistique', and 'Biochemistry'.

☐	Référence	↑	V	Titre	Processus	↑	Type de document
>				Anatomy and Histopathology team			
>				Animalerie logistique			
>				Animalerie logistique team			
>				Biochemistry			
>				Cardiology			
>				Cardiology and Respiration team			
>				Cellules ES			
>				Cognition and Behavior			



Le référencement des données et les métadonnées

Organisation, nommage et structure de données

Quelles sont les méthodes d'organisation des données ?

- ✓ Chaque projet reçoit un identifiant unique qui sera utilisé sur l'ensemble des produits
- ✓ Dans l'outil métier, dans le serveur de stockage et dans les correspondances numériques, cet identifiant unique permettra une traçabilité garantie de l'information quel que soit le stockage concerné
- ✓ Sur les outils métiers et le serveur de stockage, les données sont rangées dans des répertoires normés avec structuration établie.
- ✓ exemple d'Arborescence type pour un projet de phenotypage (voir image capture d'écran) - Les données brutes et intermédiaires sont dans le répertoire 4. Les données finales et le rapport sont dans le répertoire 5.





1_Infos-scientifiques
2_Animaux
3_Suivi-projet
4_Données
5_Rapports
9_Admin
IPXXXa TYPOLOGIE_CLIENT_GENE.txt








Quelles sont les méthodes de nommage des fichiers ?

✓ Dépend du service

3-2a0b-4815-8d1c-066e59a03d56/common-comport/COMMUN/Donnees/service/PH

Importer Créer Partager    

Partagé ave... /.common-co... /.COM... /.Don... /.ser... /.PHENOMIN-P

<input type="checkbox"/>		Nom ▲
<input type="checkbox"/>	☆	 .AppleDouble
<input type="checkbox"/>	☆	 data raw of 5609w
<input type="checkbox"/>	☆	 Worksheet
<input type="checkbox"/>	☆	 2017-02-20-Cstb-ip6487a-F-OF.xls
<input type="checkbox"/>	☆	 Stats-Cstb-ip6487a-F-OF (imported).svd



Quelle est la structure des données ?

Si les composants individuels du jeu de données sont liés les uns aux autres, décrivez succinctement.

✓ les variables/paramètres d'un jeu de données sont organisées et contextualisées, pour les mêmes individus d'un groupe d'animaux, dans un ou plusieurs tableurs, permettant de lier les paramètres d'un même item ou domaine biologique.

Si un lien existe entre différents jeux de données, décrivez succinctement

✓ Différents jeux de données ont le même identifiant "projet". il s'agit de jeux de données pour un même projet, issus de procédures différentes (ex: Echo, NIPB) et plusieurs domaines fonctionnels, et pour des individus identifiés. les procédures sont décrites dans un workflow (Cahier des charges=Phenopack) identifié.



Métadonnées

Quel type de métadonnées sont mises en œuvre ?

- ✓ Métadonnées décrivant l'objet de l'étude (décrit dans le cahier des charges)
- ✓ Métadonnées décrivant le modèle biologique (décrit dans le cahier des charges)
- ✓ Métadonnées décrivant l'échantillon biologique (décrit dans le cahier des charges et documents de saisies)
- ✓ Métadonnées décrivant l'acquisition des données (équipements et conditions d'acquisition)
- ✓ Métadonnées décrivant la donnée

Comment les métadonnées sont-elles créées, collectées et tracées, et les personnes en charges

- ✓ les métadonnées sont récoltées manuellement ou exportés automatiquement des systèmes d'acquisition.
- ✓ les métadonnées utilisées correspondent à celles reconnues dans la communauté scientifique internationale dans le domaine (notamment IMPC)
- ✓ Le choix des métadonnées relève de la responsabilité des responsables de services.
- ✓ Les données sont décrites avec un vocabulaire contrôlé



➤ Métadonnées

❖ Métadonnées [décrivant l'objet de l'étude \(décrit dans le cahier des charges\)](#) (ex: Cstb: IP6487)

Project	IMPC-Phenomin WP5		
IP	IP6487a		
CDTA line	Cstb		
Gene	Cstb		
Exact genetic background	C57BI6/N CRL		
Date of birth	51		
Diet	CD - D04		
Total number of mice	13		
Mice cohort	9F Homo		
Shipment week	6		
Upload DCC	22		2017
Age (wks)		week	year
0	Date of birth	51	2016
1			
2			
3	BREEDING CDTA		
4			
5			
6			
7	Arrivée des souris+ Tatouage Labstamp + BW	6	2017
8	dysmorpho+bw	7	2017
9	Open Field, Shirpa, Grip, Rotarod	8	2017
10	bw+acoustic startle reflex PPI, Fear conditioning	9	2017
11	energy expenditure (TSE)	10	2017
12	ECC, Echo	11	2017



➤ Métadonnées

❖ Métadonnées décrivant [l'acquisition des données \(équipements et conditions d'acquisition\)](#)

OPEN FIELD	Version	Req. Upload	Req. Analysis	Unit	Type	CENTRE SUGGESTED OPTIONS
Equipment ID IMPC_OFD_023_001	1	✓			TEXT	NI-01/557
Equipment manufacturer IMPC_OFD_024_001	1	✓	✓		TEXT	Panlab
Equipment model IMPC_OFD_025_001	1	✓	✓		TEXT	LE8822
Central zone surface area IMPC_OFD_026_001	1	✓	✓	% of total area	INT	40% of whole area
Light intensity in the centre of the arena IMPC_OFD_027_001	1	✓		Lux	FLOAT	150
Number of animals per cage IMPC_OFD_028_001	1				INT	variable
Color of arena IMPC_OFD_029_001	1				TEXT	Transparent but infrared frames
Height of the wall IMPC_OFD_030_001	1			cm	FLOAT	17
Distance from light source: if direct illuminaton IMPC_OFD_031_001	1			cm	FLOAT	not applicable
Periphery zone IMPC_OFD_032_001	1,1	✓	✓	cm	FLOAT	8
Type of analysis IMPC_OFD_033_001	1,1	✓	✓		TEXT	Wide beam break
Arena Size IMPC_OFD_034_001	1	✓			TEXT	45 X 45 cm
Experimenter ID IMPC_OFD_035_001	1,1	✓			TEXT	variable
Disinfectant IMPC_OFD_036_001	1,1				TEXT	Alcohol
Start Time IMPC_OFD_037_001	1,1				DATETIME	variable
Arena ID IMPC_OFD_038_001	1				INT	not applicable
Software Version IMPC_OFD_039_001	1	✓			TEXT	Actitrack system
Arena Material IMPC_OFD_040_001	1				TEXT	PVC
Date equipment last calibrated IMPC_OFD_041_001	1				DATETIME	before each assay



Quelles normes, ontologies, classifications, etc. sont utilisées pour décrire les données et les informations contextuelles ?

- ✓ Mammalian Phenotype Ontology (MP)
- ✓ MIMPP (Minimal Information for Mouse Phenotyping Procedures)
- ✓ European Nucleotide Archive (ENA)
- ✓ Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE)
- ✓ The Digital MIQE Guidelines Update: Minimum Information for Publication of Quantitative Digital PCR



Partage et réutilisation

Quels sont les produits de la recherche concernés par cette question

✓ Selon le financeur du projet, les données peuvent être soumises à des périodes d'embargo et leur propriété peut être entièrement celle du partenaire. Dans le cas d'un partenaire public, nous encourageons la diffusion des données dès que possible, dans un principe FAIR.

Quelles sont les actions menées pour permettre la re-création des données à l'identique ou à l'équivalent ?

- ✓ Utilisation d'un protocole standardisé de collecte des informations
- ✓ Tenue d'une traçabilité par l'opérateur (e.g. cahier de laboratoire, notebook, etc...)
- ✓ Utilisation de standards dans la communauté (outils, technologies, formats, etc...)



Gestion des données chaudes et tièdes

Estimation de la taille des données pour les produits listés

Listez ici le(les) produit(s) de la recherche concerné(s) par la question

✓ les produits relatifs à un projet contenant à la fois la création et la caractérisation d'un modèle rongeur

Quelle est la taille estimée (en KB, MB, GB, TB)

✓ 5 MB-500 GB environ

Conditions de stockage pendant le projet

✓ Les données brutes numériques et textuelles sont stockées dans les machines d'acquisition/postes de travail et serveurs de l'IGBMC

✓ Les données lourdes (images, vidéos) ou ne disposant pas de trames de saisie dans les bases internes sont stockées dans des serveurs de l'IGBMC

Où les données sont-elles stockées ?

✓ Machine d'acquisition/poste de travail

✓ Serveurs IGBMC

✓ Base de données interne (Biox, Goliath,...)

✓ Disque dur externe



Implémentation de la sécurité des données

Sauvegarde des données

Quelles sont les stratégies de sauvegarde employées ?

- ✓ Les données brutes sont sauvegardées dans les machines d'acquisition, puis dans transférées dans les bases internes et/ou dans les serveurs de l'IGBMC.
- ✓ Les données analysées sont sauvegardées dans les bases internes et/ou serveurs de l'IGBMC (qui sont sur site et en externe)

Quelle est la fréquence des sauvegardes ?

- ✓ 2 fois par jour pendant 1 semaine
- ✓ Chaque semaine, pendant 1 mois
- ✓ Jusqu'à transfert **au client ou archivage**

Quelle est la durée de conservation des sauvegardes ?

- ✓ Les données sont sauvegardées 1 an après transfert au donneur d'ordre

Quelles sont la/les personne(s)/entité(s) responsable(s) des sauvegardes ?

- ✓ Responsable et membres de services
- ✓ Responsable de service informatique IGBMC pour les serveurs



Sécurité des accès aux données

Mesures ou dispositions mises en place pour garantir la sécurité des accès aux données ?

- ✓ les postes de travail sont accessibles uniquement avec des identifiants/mots de passes personnels
- ✓ les répertoires partagés contenant les données des projets sont partagés uniquement avec le personnel des services et des départements concernés
- ✓ les bases de données sont aussi accessibles uniquement sur le réseau interne, sur un poste appartenant au parc informatique avec un identifiant / mot de passe
- ✓ sur certains postes d'acquisition liés à des machines plus anciennes, les données brutes sont stockées avant leur transfert. Les postes concernés sont dans des zones d'accès contrôlés avec badge individuel.

Quelles sont la/les personne(s)/entité(s) responsable(s) de la sécurité des accès ?

- ✓ Responsable de service
- ✓ Responsable de service informatique
- ✓ Gestion des mots de passe et des badges centralisés pour les bâtiments IGBMC/ICS



Partage de données pendant le projet

- ✓ En interne pour les différents supports de stockage
- ✓ Limité en externe, avec donneur d'ordre/client (par messagerie électronique, par partage sécurisé via le serveur de fichier avec date d'expiration et mot de passe individuel, par disque dur) ; selon la demande du partenaire.

Quel est le devenir de ces données ?

- ✓ Elles vont être transférées au donneur d'ordre

Si pertinent, décrivez les modalités de transfert

Selon la demande du partenaire :

- ✓ Par messagerie électronique
- ✓ Par partage sécurisé via le serveur de fichier avec date d'expiration et mot de passe individuel
- ✓ Par disque dur

- ✓ les données vont être gardées pendant 1 an après transmission au donneur d'ordre



Exemple de données IMPC



Outre ses activités de service, l'ICS est impliqué dans plusieurs programmes de recherche.

❖

❖ **IMPC (international Mouse Phenotyping Consortium)**

❖



Objectif/Défi IMPC

Construire le premier catalogue complet des fonctions du génome d'un mammifère, la souris



IMPC – Un effort global à travers le monde

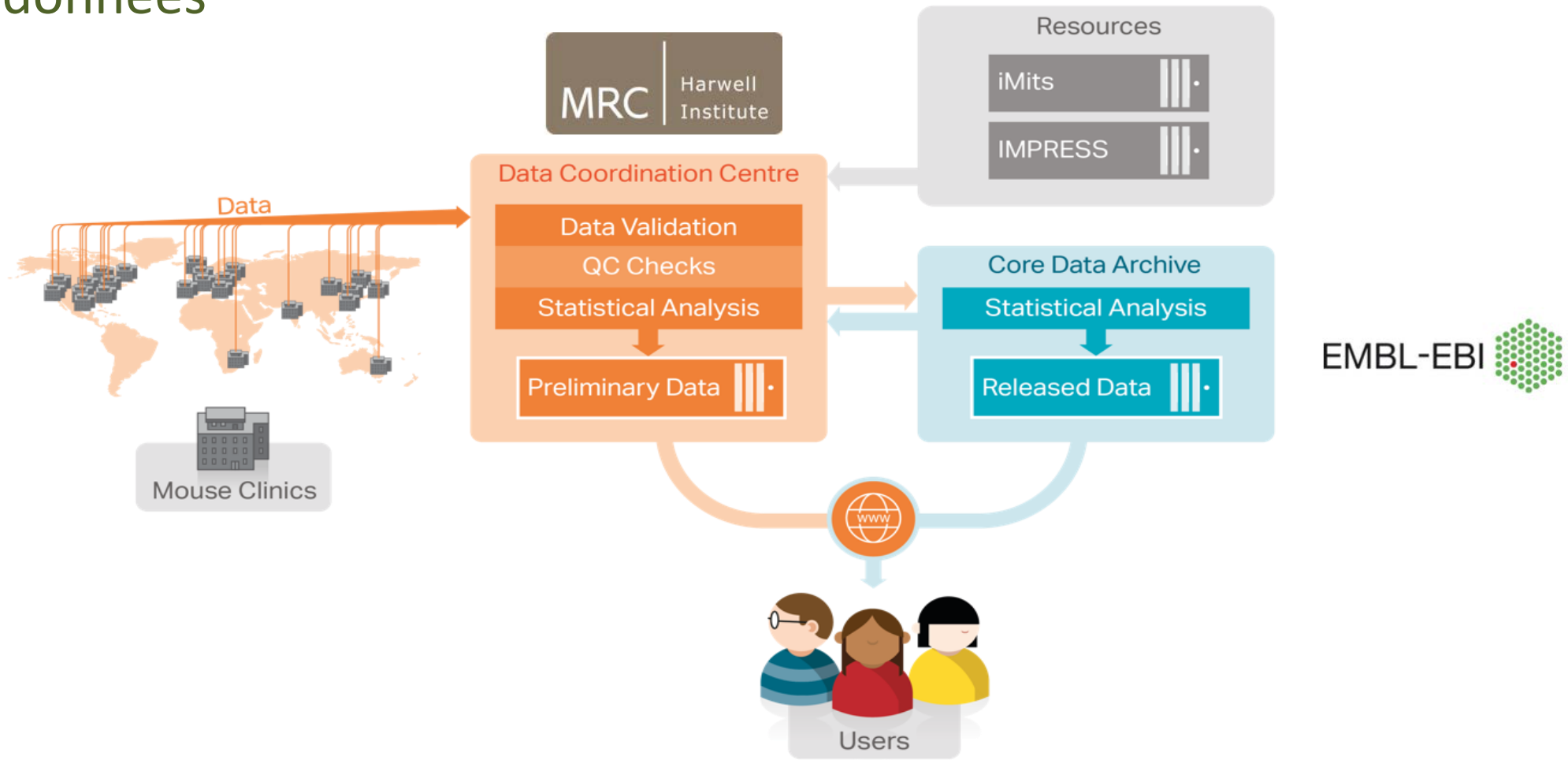




IMPC – En pratique

- Générer une souris KO pour chaque gène codant
- Phénotypage complet de chaque mutant à l'aide de pipelines standardisés
- Toutes les données générées par les cliniques de souris du monde entier
 - ❖ Transférées au centre de coordination des données (DCC) à Harwell pour la validation des données et les contrôles de qualité
 - ❖ Puis déposés dans un centre d'archivage (EBI, Cambridge) où elles sont analysées statistiquement et mises à la disposition de la communauté

Flux de données





6. Gestion des données froides

1. Listez ici le(les) produit(s) de la recherche concerné(s) par la question

✓ TOUS|

2. Y-a-t-il des restrictions gouvernant le partage de ces données ?

✓ Non

3. Quel est le devenir de ces données ?

✓ Elles vont être stockées et conservées à long terme

3.b.1. Quels sont les critères et/ou les règles de sélection des données à archiver ?

- ✓ Données de projets de recherche publiques
- ✓ Les données sont utilisées dans une publication
- ✓ Les données vont ou peuvent être réutilisées dans des projets ultérieurs

3.b.2. Y-a-t'il d'autres critères de selection que ceux listés ci-dessus ?

✓ - Contribution à la Science ouverte - contribution aux 3Rs



3.b.3. Qui (personne ou entité) est responsable de la sélection des données à archiver ?

- ✓ Responsables des services; discuté et agréés par des groupes de travail

3.b.5. Quelle est la date d'archivage ?

- ✓ Dès la validation des résultats d'analyse

3.b.7. Où les données seront-elles stockées ?

1. Indiquez l'URL

- ✓ <http://www.mousephenotype.org/>

2. L'entrepôt est-il certifié ? Si oui, indiquez la certification

- ✓ Oui

3. Avez vous des commentaires à faire au sujet de l'entrepôt ?

- ✓ Licence Apache 2.0
-



3.b.8. Où seront stockées les métadonnées ?

- ✓ Les métadonnées sont stockées au même endroit que les données

3.b.10. Les données seront-elles partagées ?

- ✓ Oui

3.b.10.b.1. Quelles sont les modalités de partage ?

- ✓ En externe pour tout le monde

3.b.10.b.4. À partir de quand les données seront-elles accessibles ?

- ✓ Après curation et validation

3.b.11. Si pertinent, quelle(s) licence(s) seront attribuées aux données ?

- ✓ CC-BY-4.0

3.b.12. Quelle est la durée de stockage envisagée ?

- ✓ NA



7. Questions autour du cadre légal et éthique

2. Quel est le cadre juridique applicable ?

✓ La réglementation européenne

3. Ces données contiennent-elles des données personnelles ?

✓ Non

4. Ces données contiennent-elles des données sensibles ?

✓ Non

5. Quels sont les droits de la propriété intellectuelle ou industrielle applicables pour ces données ?

▲ 1. Les données sont-elles protégées par des droits de propriété intellectuelle ou industrielle ?

✓ Non



Portail IMPC: <http://www.mousephenotype.org/>

Home | IMPC | International M... x +

mousephenotype.org

Data release 20.0 is now here! With a total 8,707 knockout genes phenotyped. Read the **full report here**

GENES PHENOTYPES HELP, NEWS, BLOG

Search All 8707 Knockout Data...

What you need to know about IMPC data

What you need to know about IMPC data

- » We generate our own data, it is **not aggregated** from publications
- » We are targeting genes that are *yet to be studied*
- » We have **standardised** production methods to knock out genes

8707 total knockout genes phenotyped

- » Number of phenotyped lines: 9393
- » Statistically Significant Calls: 104530
- » Data Release Version: 20.0
- » Published: 21 November 2023

This site uses cookies. Some of them are essentials while other help us improve your experience. [Learn more](#)

Essential Analytics Close Save preferences

Taper ici pour 10:20 05/02/2024



Conclusion

- Un PGD entité/structure déploie une architecture de questions pour contextualiser les données générées dans une structure ou dans une discipline (e.g. Bio-imagerie) et utilisant les ressources de cette entité.
- Un PGD projet peut utiliser la trame d'un PGD entité, et répond aux questions prédéfinies
 - ❖ Permettant de décrire les éléments du cycle de vie des données, les rendant FAIR
 - ❖ Facilitant la reproductibilité des données et améliorant l'efficacité de recherche
 - ❖ Contribuant aux principes des 3Rs

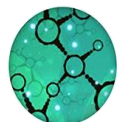


THANK YOUs....

Anne-Francoise Adam-Blondon, Christophe Antoniewski, **Faisal Bekkouche**, David Benaben, Audrey Bihouee, Catherine Blanc, Sidina Boudaakat, **Christophe Bruley**, Julie Cazareth, Nicole Charrière, Stéphanie Cheviron, Noémi Cobolet, Sophie Desset, Gaëlle Gauvrit, Guillaume Gay, Virginie Georget, Franck Giacomoni, Nadia Goué, Thomas Guilbert, Christine Jacquemot, Suzanne Lauriou, Hamid Meziane, Sylvain Milanese, Cathleen Mirande-Ney, **Perrine Paul-Gilloteaux**, Mélanie Petera, Nadia Ponts, Marina Popleteeva, Sylvain Pringent, Angela Saenz, **Jean Salamero**, Fanny Sebire, **Marek Suchánek**, Raluca Teusan, Julie Thompson, Joseph Tran, Jérôme Vialaret, Valérie Warth

the teams of DSW and DMP-OPIDoR,  **DSW** 

... and all the others we might have missed to mention...



FRANCE-BIOIMAGING



FRANCE
GENOMIQUE



ProFI
PROTEOMICS



METABOHUB



INRAE



Science ouverte

Université de Strasbourg

