



HAL
open science

DMP projet ANR DIG-AI

Jérôme Azé, Antony Champion, Konstantin Todorov, Alexandre Grondin,
Mikaël Lucas, François Scharffe, Pascal Gantet, Pierre Larmande

► **To cite this version:**

Jérôme Azé, Antony Champion, Konstantin Todorov, Alexandre Grondin, Mikaël Lucas, et al.. DMP projet ANR DIG-AI. IRD. Institut de Recherche pour le Développement. 2023. hal-04161144

HAL Id: hal-04161144

<https://hal.science/hal-04161144>

Submitted on 27 Jul 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

DMP du projet "ANR DIG-AI"

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

Renseignements sur le plan

Titre du plan	DMP du projet "ANR DIG-AI"
Livable	D0.2
Version	Version initiale
Objet/périmètre du plan	Description des logiciels et datasets disponibles et utilisés en debut du projet ainsi que ceux qui seront produits dans le cadre du projet
Domaines de recherche (selon classification de l'OCDE)	Computer and information sciences
Langue	fra
Date de création	2023-07-13
Date de dernière modification	2023-07-27
Identifiant	https://hal.science/hal-04161144
Type d'identifiant	DOI
Licence	Creative Commons Attribution 4.0 International

Documents

(publications, rapports, brevets, plan expérimental....), sites web associés

- AgroLD Web application : https://github.com/SouthGreenPlatform/AgroLD_webapp
- AgroLD_ETL : https://github.com/SouthGreenPlatform/AgroLD_ETL
- DLINKER :
- BioGruT :
- AgroLD Datasets : <https://zenodo.org/record/4694518>
- OryzaGP : <https://github.com/pierrelarmande/OryzaGP>
- AgroLD_Data_Linking_Benchmark : https://github.com/pierrelarmande/AgroLD_Data_Linking_Benchmark
- AgroLD Augmentation benchmark dataset :
- Anatomics :
- IonoMil :
- AgronoMil :
- PlastiMil :
- TDCor : <https://cran.r-project.org/web/packages/TDCor/index.html>
- PANTHEON :
- LR Dataset :

Renseignements sur le projet

Titre du projet	Plan de gestion du projet ANR DIG-AI
Acronyme	DIG-AI
Résumé	<p>Afin de relever le défi de la forte demande alimentaire dans les pays du SUD dans un contexte de changement climatique, une meilleure compréhension des relations génotype-phénotype est importante pour améliorer les capacités de production des cultures. Les enjeux consistent à intégrer des informations de différents niveaux dans un modèle global par une approche systémique afin de comprendre le fonctionnement réel du système biologique. L'objectif principal de DIG-AI est de développer des méthodes de science des données pour tirer parti des connaissances exploitables des données biologiques et identifier les gènes candidats du monde réel pour l'amélioration des plantes. Nous nous concentrerons sur les objectifs spécifiques suivants : 1) Exploiter les méthodes d'ingénierie des connaissances grâce à l'utilisation d'ontologies pour formuler des hypothèses de recherche liant génotype au phénotype, 2) Gérer efficacement de grands volumes de données biologiques pour extraire des connaissances, 3) Utiliser la base de connaissances AgroLD construire des réseaux d'interactions moléculaires entre gènes et phénotypes à partir de données éparses (articles scientifiques, bases de données publiques, données expérimentales), 4) Identifier des gènes clés pour l'amélioration des plantes parmi des centaines de résultats potentiels.</p>

Sources de financement

- Agence Nationale de la Recherche : ANR-22-CE23-0012

Date de début 2023-01-05

Date de fin 2026-06-30

Partenaires

- Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier (199111950H)
- Diversité, Adaptation et Développement des plantes (201119426X)

Produits de recherche :

1. AgroLD Web application (Logiciel)
2. AgroLD_ETL
3. DLINKER
4. Biological Gated Recurrent Unit Transformer (Logiciel)
5. AgroLD_datasets (Jeu de données)
6. OryzaGP (Jeu de données)
7. AgroLD Data Linking Benchmark (Jeu de données)
8. AgroLD Augmentation Dataset Benchmark
9. Anatomics (Jeu de données)
10. IonoMil (Jeu de données)
11. AgronoMil (Jeu de données)
12. PlastiMil (Jeu de données)
13. Time Delay Correlation algorithm (Logiciel)
14. PANTHEON (Logiciel)
15. Lateral root dataset (Jeu de données)
16. Time-series transcript profiling analysis in stem base of rice crown rootless 1 mutant after ectopic expression induction by dexamethasone of the CRL1 gene (Jeu de données)

Contributeurs

Nom	Affiliation	Rôles
Aze Jerome - https://orcid.org/0000-0002-7372-729X	Universite Montpellier	
Champion Antony - https://orcid.org/0000-0002-3731-1510	IRD	
Gantet Pascal - https://orcid.org/0000-0003-1314-0187	Universite Montpellier	<ul style="list-style-type: none"> • Personne contact pour les données (TS_CR_Rice) • Responsable du dépôt et de la diffusion des données (TS_CR_Rice)
GRONDIN Alexandre - 0000-0001-6726-6274	IRD	<ul style="list-style-type: none"> • Personne contact pour les données (AgronoMil, Anatomics, IonoMil, PlastiMil)
LARMANDE PIERRE		<ul style="list-style-type: none"> • Coordinateur de projet • Personne contact pour les données (AgroLD_webapp, AgroLD_DL, AgroLD_ETL, BioGruT, DLINKER, AgroLD_datasets, OryzaGP, AgroLD_AUG) • Responsable de la documentation des données (AgroLD_datasets) • Responsable de la qualité des données (OryzaGP) • Responsable du plan • Responsable du stockage des données (AgroLD_datasets) • Responsable du traitement et de l'analyse des données (OryzaGP)
Lucas Mikael - https://orcid.org/0000-0001-6286-7759	IRD	<ul style="list-style-type: none"> • Personne contact pour les données (PANTHEON, LR Dataset, TDCor)
Scharffe Francois - https://orcid.org/0000-0002-0010-0058	Universite Montpellier	
Todorov Konstantin - https://orcid.org/0000-0002-9116-6692	Universite Montpellier	

DMP du projet "ANR DIG-AI"

1. Description des données et collecte ou réutilisation de données existantes

AgroLD Web application

1.1 Description générale du produit de recherche

Nom	AgroLD Web application
Description	Web application of the AgroLD platform
Type	Logiciel
Workpackage	WP1, WP2, WP3
Mots clés (texte libre)	
Langue	eng
Date de publication	2016-06-06
Identifiant pérenne	https://doi.org/10.1371/journal.pone.0198270
Type d'identifiant	DOI
Contient des données personnelles ?	Non
Contient des données sensibles ?	Non
Prend en compte des aspects éthiques ?	Non

1.2 Est-ce que des données existantes seront réutilisées ?

Justification	AgroLD application will serve a basis for the project
----------------------	---

1.3 Comment seront produites/collectées les nouvelles données ?

Nom de la méthode	AgroLD_ETL
Description	We will use and extend this method to produce new datasets

AgroLD_ETL

1.1 Description générale du produit de recherche

Nom	AgroLD_ETL
------------	------------

1.2 Est-ce que des données existantes seront réutilisées ?

1.3 Comment seront produites/collectées les nouvelles données ?

Nom de la méthode	AgroLD_ETL
Description	Python script and RDF models will be used to produce the data

DLINKER

1.1 Description générale du produit de recherche

Nom	DLINKER
Description	Data Linking Software using knowledge graph
Workpackage	WP2
Mots clés (texte libre)	

1.2 Est-ce que des données existantes seront réutilisées ?

Justification	yes Dlinker software will be open source
----------------------	--

1.3 Comment seront produites/collectées les nouvelles données ?

Description	non applicable
--------------------	----------------

Biological Gated Recurrent Unit Transformer

1.1 Description générale du produit de recherche

Nom	Biological Gated Recurrent Unit Transformer
Description	Software developed to detect biological entities in scientific articles and texts
Type	Logiciel
Workpackage	WP1, WP2, WP3
Mots clés (texte libre)	

1.2 Est-ce que des données existantes seront réutilisées ?

Justification	BioGrut wil be share through the GitHub repository https://github.com/ANR-DIG-AI
1.3 Comment seront produites/collectées les nouvelles données ?	
Description	non applicable

AgroLD_datasets	
1.1 Description générale du produit de recherche	
Nom	AgroLD_datasets
Description	RDF Datasets produced for the AgroLD platform
Type	Jeu de données
Workpackage	WP1, WP2, WP3
Mots clés (texte libre)	RDF datasets
Date de publication	2021-01-01
Identifiant pérenne	https://zenodo.org/record/4694518
Type d'identifiant	DOI
Contient des données personnelles ?	Non
Contient des données sensibles ?	Non
Prend en compte des aspects éthiques ?	Non
1.2 Est-ce que des données existantes seront réutilisées ?	
Justification	oui
1.3 Comment seront produites/collectées les nouvelles données ?	
Nom de la méthode	AgroLD_ETL
Nature des données	Modèle

OryzaGP	
1.1 Description générale du produit de recherche	

Nom	OryzaGP
Description	Gene / Protein Named entity recognition benchmark datasets
Type	Jeu de données
Workpackage	WP1
Mots clés (texte libre)	NER, Benchmark
Langue	eng
Date de publication	2021-04-01
Identifiant pérenne	https://doi.org/10.5808/GI.2019.17.2.e17
Type d'identifiant	DOI
Contient des données personnelles ?	Non
Contient des données sensibles ?	Non
Prend en compte des aspects éthiques ?	Non

1.2 Est-ce que des données existantes seront réutilisées ?

Justification yes

1.3 Comment seront produites/collectées les nouvelles données ?

Nom de la méthode BioGruT
Description BioGrut is a NER software
Nature des données Modèle

AgroLD Data Linking Benchmark

1.1 Description générale du produit de recherche

Nom	AgroLD Data Linking Benchmark
Description	AgroLD Data Linking Benchmark aims as providing annotated dataset for training and validating Data Linking softwares
Type	Jeu de données
Workpackage	WP2
Mots clés (texte libre)	Data linking
Langue	eng
Date de publication	2022-11-10
Identifiant pérenne	https://github.com/pierrelarmande/AgroLD_Data_Linking_Benchmark
Type d'identifiant	URL
Contient des données personnelles ?	Non
Contient des données sensibles ?	Non
Prend en compte des aspects éthiques ?	Non

1.2 Est-ce que des données existantes seront réutilisées ?

Justification yes the dataset will be reuse in the frame of WP2

1.3 Comment seront produites/collectées les nouvelles données ?

Nom de la méthode DLINKER
Nature des données Modèle

AgroLD Augmentation Dataset Benchmark

1.1 Description générale du produit de recherche

Nom	AgroLD Augmentation Dataset Benchmark
Description	This benchmark is dedicated to the evaluation of Augmentation software
Workpackage	WP2
Mots clés (texte libre)	Augmentation
Langue	eng
Date de publication	2023-10-01
Identifiant pérenne	AgroLD-AUGM
Type d'identifiant	Identifiant local
Contient des données personnelles ?	Non
Contient des données sensibles ?	Non
Prend en compte des aspects éthiques ?	Non

1.2 Est-ce que des données existantes seront réutilisées ?	
Justification	Yes they will be use by the WP2
1.3 Comment seront produites/collectées les nouvelles données ?	
Description	manually

Anatomics	
1.1 Description générale du produit de recherche	
Nom	Anatomics
Description	drough tolerance datasets for root architecture
Type	Jeu de données
Workpackage	WP1, WP4
Mots clés (texte libre)	
Langue	eng
1.2 Est-ce que des données existantes seront réutilisées ?	
Justification	yes datasets will be reused in other WP
1.3 Comment seront produites/collectées les nouvelles données ?	

Description Data will be collected in another context / Project

IonoMil

1.1 Description générale du produit de recherche

Nom IonoMil
Description Experimental dataset for the Drought Stress use case
Type Jeu de données
Workpackage WP1,WP4
Mots clés (texte libre)
Langue eng
Identifiant pérenne IonoMil
Type d'identifiant Identifiant local
Contient des données personnelles ? Non
Contient des données sensibles ? Non
Prend en compte des aspects éthiques ? Oui

1.2 Est-ce que des données existantes seront réutilisées ?

Justification yes the dataset will be shared and reuse

1.3 Comment seront produites/collectées les nouvelles données ?

Nature des données Données expérimentales

AgronoMil

1.1 Description générale du produit de recherche

Nom	AgronoMil
Description	Agronomy dataset for pearl millet
Type	Jeu de données
Workpackage	WP1,WP4
Mots clés (texte libre)	
Langue	eng
Identifiant pérenne	Agronomil
Type d'identifiant	Identifiant local
Contient des données personnelles ?	Non
Contient des données sensibles ?	Non
Prend en compte des aspects éthiques ?	Oui

1.2 Est-ce que des données existantes seront réutilisées ?

Justification Yes datasets will be reused

1.3 Comment seront produites/collectées les nouvelles données ?

PlastiMil

1.1 Description générale du produit de recherche

Nom	PlastiMil
Description	Experimental dataset on root plasticity observation for Pearl millet
Type	Jeu de données
Workpackage	WP1, WP4
Mots clés (texte libre)	
Langue	eng
Identifiant pérenne	Plastimil
Type d'identifiant	Identifiant local
Contient des données personnelles ?	Non
Contient des données sensibles ?	Non
Prend en compte des aspects éthiques ?	Oui

1.2 Est-ce que des données existantes seront réutilisées ?

Justification Yes the data will be reused in the frame of the project and outside

1.3 Comment seront produites/collectées les nouvelles données ?

Time Delay Correlation algorithm

1.1 Description générale du produit de recherche

Nom Time Delay Correlation algorithm

Description Bioinformatics software to infer gene regulation interaction from experimental transcriptomics time series datasets

Type Logiciel

Workpackage WP1, WP4

Mots clés

- (inference)
- (algorithm)
- (transcriptomics)
- (gene network)

Mots clés (texte libre)

Langue eng

Identifiant pérenne TdCor

Type d'identifiant Identifiant local

Contient des données personnelles ? Non

Contient des données sensibles ? Non

Prend en compte des aspects éthiques ? Oui

1.2 Est-ce que des données existantes seront réutilisées ?

Justification the software is open source

1.3 Comment seront produites/collectées les nouvelles données ?

Description We will use TdCor to produce new data

Nature des données Modèle

PANTHEON

1.1 Description générale du produit de recherche

Nom	PANTHEON
Description	Pantheon is a software package created to allow dynamical modeling of network behavior. It has been validated using experimental data for regulatory genes networks active in plant root development.
Type	Logiciel
Workpackage	WP1, WP4
Mots clés (texte libre)	
Langue	eng
Identifiant pérenne	Pantheon
Type d'identifiant	Identifiant local
Contient des données personnelles ?	Non
Contient des données sensibles ?	Non
Prend en compte des aspects éthiques ?	Oui

1.2 Est-ce que des données existantes seront réutilisées ?

Justification Pantheon is open source and thus will be reuse

1.3 Comment seront produites/collectées les nouvelles données ?

Description we wil produce new data according the datasets used in the project

Nature des données Modèle

Lateral root dataset

1.1 Description générale du produit de recherche

Nom	Lateral root dataset
Description	The dataset is built to observe gene network regulation across lateral root development. It has been generated from dense long term time series transcriptomics experimental data.
Type	Jeu de données
Workpackage	WP1,WP4
Mots clés (texte libre)	
Langue	eng
Identifiant pérenne	LRDataset
Type d'identifiant	Identifiant local
Contient des données personnelles ?	Non
Contient des données sensibles ?	Non
Prend en compte des aspects éthiques ?	Oui

1.2 Est-ce que des données existantes seront réutilisées ?

Justification Yes the dataset will be reused

1.3 Comment seront produites/collectées les nouvelles données ?

Description Methods have been described in the publication

Time-series transcript profiling analysis in stem base of rice crown rootless 1 mutant after ectopic expression induction by dexamethasone of the CRL1 gene

1.1 Description générale du produit de recherche

Nom	Time-series transcript profiling analysis in stem base of rice crown rootless 1 mutant after ectopic expression induction by dexamethasone of the CRL1 gene
Description	<p>Lateral Organ Boundary Domain (LBD) transcription factors are specific of plants and are involved in the control of development. One LBD clade is related to the control of root development (Coudert et al., 2013, Mol. Biol. Evol. 30, 569-572). Belonging to this clade, CROWN ROOTLESS 1 controls the initiation of crown roots in rice (Inukai Plant Cell, 17, 1387-1396, Liu et al., 2005, Plant J., 43, 47-56). The aim of this study was to identify the genes that are regulated by CRL1 in a time-dependant manner after its induction. A total of 1748 differentially expressed genes (DEG) were identified, 1034 of them upregulated, and 706 downregulated, and 14 both upregulated and downregulated at some point during the time series. Data analysis revealed three distinct phases in the time series that can represent three developemntal stage during crown root primordia formation.</p> <p>Affymetrix microarrays were processed in the Microarray Core Facility "Transcriptome" of the Institute in Regenerative Medicine and Biotherapy, CHU de Montpellier-INSERM-UM Montpellier, http://irmb.chu-montpellier.fr/ .</p>
Type	Jeu de données
Workpackage	WP1,WP4
Mots clés (texte libre)	
Langue	eng
Date de publication	2020-08-27
Identifiant pérenne	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE123819
Type d'identifiant	URL
Contient des données personnelles ?	Non
Contient des données sensibles ?	Non
Prend en compte des aspects éthiques ?	Oui
1.2 Est-ce que des données existantes seront réutilisées ?	
Justification	yes dataset is public
1.3 Comment seront produites/collectées les nouvelles données ?	
Nom de la méthode	sequencins
Nature des données	Données expérimentales

2. Documentation et qualité des données

AgroLD Web application	
2.1 Quelles métadonnées et quelle documentation (par exemple mode d'organisation des données) accompagneront les données ?	
Description	we will follow the metadata described in the following link https://github.com/SouthGreenPlatform/AgroLD_ETL/tree/master/model

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

Description we will develop python scripts and schacl scripts to ensure data quality

AgroLD_ETL

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

Description we will use the following metadata schema to produce data
https://github.com/SouthGreenPlatform/AgroLD_ETL/tree/master/model

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

Description we will use python and schacl scripts to ensure data quality and consistency

DLINKER

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

Description Metadata and documentation will be available through the GitHub <https://github.com/ANR-DIG-AI>

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

Description non applicable

Biological Gated Recurrent Unit Transformer

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

Description Metadata and Documentation will be available through the GitHub repository <https://github.com/ANR-DIG-AI>

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

Description code review and passing tests

AgroLD_datasets

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

Description Datasets will be described with the following metadata
https://github.com/SouthGreenPlatform/AgroLD_ETL/tree/master/model

Standards de métadonnées/données

- Dublin Core : <https://rdamsc.bath.ac.uk/msc/m15>
- PROV : <https://rdamsc.bath.ac.uk/msc/m33>
- schema.org : <https://rdamsc.bath.ac.uk/msc/m101>

Code langue des métadonnées eng

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

Description python scripts and schacl scripts

OryzaGP

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

Description Documentation will be attached at the GitHub repository <https://github.com/ANR-DIG-AI>

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

Description We will follow the classical machine learning approaches to ensure a good accuracy of the models and validate results with our collaborators

AgroLD Data Linking Benchmark

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

Description Metadata will be available on the Github repository

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

Question sans réponse.

AgroLD Augmentation Dataset Benchmark

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

Description Documentation will be attached on the Github page

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

Description Dataset will be generated manually by expert

Anatomics

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

Description MINIMAL metadata will be attached with the datasets - they will follow requierment from the dataverse repository

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

Description State of art methods will be used to produce the data

IonoMil

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

Description not defined yet

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

Description not defined yet

AgronoMil

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

Description not defined yet

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

Description not defined yet

PlastiMil

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

Description not defined yet

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

Description not defined yet

Time Delay Correlation algorithm

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

Description Documentation will be available through the software GitHub page

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

Description Passing test are define to ensure quality

PANTHEON

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

Description documentation will be available through the GitHub page

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

Description Test passing will be set up to ensure data quality

Lateral root dataset

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

Description not defined yet

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

Description human expert will validate data quality

Time-series transcript profiling analysis in stem base of rice crown rootless 1 mutant after ectopic expression induction by dexamethasone of the CRL1 gene

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

Description standard metadata provided by the platform

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

Description already checked

3. Exigences légales et éthiques, code de conduite

AgroLD Web application

3.1 Quelles seront les mesures appliquées pour assurer la protection des données à caractère personnel ?

Description non applicable

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

Description non applicable

3.3 Quels sont les aspects éthiques à prendre en compte lors de la collecte des données ?

Description	MTA and DTA if necessary
-------------	--------------------------

AgroLD_ETL	
3.1 Quelles seront les mesures appliquées pour assurer la protection des données à caractère personnel ?	
Description	non applicable
3.2 Comment les autres questions juridiques, comme la titularité ou les droits de propriété intellectuelle sur les données, seront-elles abordées ? Quelle est la législation applicable en la matière ?	
Description	non applicable
3.3 Quels sont les aspects éthiques à prendre en compte lors de la collecte des données ?	
Description	non applicable

DLINKER	
3.1 Quelles seront les mesures appliquées pour assurer la protection des données à caractère personnel ?	
Description	non applicable
3.2 Comment les autres questions juridiques, comme la titularité ou les droits de propriété intellectuelle sur les données, seront-elles abordées ? Quelle est la législation applicable en la matière ?	
Description	non applicable
3.3 Quels sont les aspects éthiques à prendre en compte lors de la collecte des données ?	
Description	non applicable

Biological Gated Recurrent Unit Transformer	
3.1 Quelles seront les mesures appliquées pour assurer la protection des données à caractère personnel ?	

Description	non applicable
3.2 Comment les autres questions juridiques, comme la titularité ou les droits de propriété intellectuelle sur les données, seront-elles abordées ? Quelle est la législation applicable en la matière ?	
Description	non applicable
3.3 Quels sont les aspects éthiques à prendre en compte lors de la collecte des données ?	
Description	non applicable

AgroLD_datasets	
3.1 Quelles seront les mesures appliquées pour assurer la protection des données à caractère personnel ?	
Description	non applicable
3.2 Comment les autres questions juridiques, comme la titularité ou les droits de propriété intellectuelle sur les données, seront-elles abordées ? Quelle est la législation applicable en la matière ?	
Description	non applicable
3.3 Quels sont les aspects éthiques à prendre en compte lors de la collecte des données ?	
Description	non applicable

OryzaGP	
3.1 Quelles seront les mesures appliquées pour assurer la protection des données à caractère personnel ?	
Description	non applicable
3.2 Comment les autres questions juridiques, comme la titularité ou les droits de propriété intellectuelle sur les données, seront-elles abordées ? Quelle est la législation applicable en la matière ?	
Description	non applicable

3.3 Quels sont les aspects éthiques à prendre en compte lors de la collecte des données ?

Description non applicable

AgroLD Data Linking Benchmark

3.1 Quelles seront les mesures appliquées pour assurer la protection des données à caractère personnel ?

Description non applicable

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

Description non applicable

3.3 Quels sont les aspects éthiques à prendre en compte lors de la collecte des données ?

Description non applicable

AgroLD Augmentation Dataset Benchmark

3.1 Quelles seront les mesures appliquées pour assurer la protection des données à caractère personnel ?

Description non applicable

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

Description non applicable

3.3 Quels sont les aspects éthiques à prendre en compte lors de la collecte des données ?

Description non applicable

Anatomics

3.1 Quelles seront les mesures appliquées pour assurer la protection des données à caractère personnel ?

Description non applicable

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

Description non applicable

3.3 Quels sont les aspects éthiques à prendre en compte lors de la collecte des données ?

Description non applicable

IonoMil

3.1 Quelles seront les mesures appliquées pour assurer la protection des données à caractère personnel ?

Description non applicable

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

3.3 Quels sont les aspects éthiques à prendre en compte lors de la collecte des données ?

Question sans réponse.

AgronoMil

3.1 Quelles seront les mesures appliquées pour assurer la protection des données à caractère personnel ?

Description non applicable

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

3.3 Quels sont les aspects éthiques à prendre en compte lors de la collecte des données ?

Question sans réponse.

PlastiMil

3.1 Quelles seront les mesures appliquées pour assurer la protection des données à caractère personnel ?

Description	non applicable
-------------	----------------

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

3.3 Quels sont les aspects éthiques à prendre en compte lors de la collecte des données ?

Question sans réponse.

Time Delay Correlation algorithm

3.1 Quelles seront les mesures appliquées pour assurer la protection des données à caractère personnel ?

Description	non applicable
-------------	----------------

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

Question sans réponse.

3.3 Quels sont les aspects éthiques à prendre en compte lors de la collecte des données ?

Question sans réponse.

PANTHEON

3.1 Quelles seront les mesures appliquées pour assurer la protection des données à caractère personnel ?

Description non applicable

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

Question sans réponse.

3.3 Quels sont les aspects éthiques à prendre en compte lors de la collecte des données ?

Question sans réponse.

Lateral root dataset

3.1 Quelles seront les mesures appliquées pour assurer la protection des données à caractère personnel ?

Description non applicable

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

Question sans réponse.

3.3 Quels sont les aspects éthiques à prendre en compte lors de la collecte des données ?

Question sans réponse.

Time-series transcript profiling analysis in stem base of rice crown rootless 1 mutant after ectopic expression induction by dexamethasone of the CRL1 gene

3.1 Quelles seront les mesures appliquées pour assurer la protection des données à caractère personnel ?

Description non applicable

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

Description non applicable

3.3 Quels sont les aspects éthiques à prendre en compte lors de la collecte des données ?

Description non applicable

4. Traitement et analyse des données

AgroLD Web application

4.1 Comment et avec quels moyens seront traitées les données ?

Description We will use AgroLD-ETL

AgroLD_ETL

4.1 Comment et avec quels moyens seront traitées les données ?

Question sans réponse.

DLINKER

4.1 Comment et avec quels moyens seront traitées les données ?

Description non applicable

Biological Gated Recurrent Unit Transformer

4.1 Comment et avec quels moyens seront traitées les données ?

Description non applicable

AgroLD_datasets

4.1 Comment et avec quels moyens seront traitées les données ?

Description we will use AgroLD_ETL to create the datasets

Equipements, plateaux techniques

- South Green : <https://cat.opidor.fr/index.php/South Green>

OryzaGP

4.1 Comment et avec quels moyens seront traitées les données ?

Description Data will be treated with BioGrut

Equipements, plateaux techniques

- IDRIS : <https://cat.opidor.fr/index.php/IDRIS>
- South Green : <https://cat.opidor.fr/index.php/South Green>

AgroLD Data Linking Benchmark

4.1 Comment et avec quels moyens seront traitées les données ?

Description	We will use DLINKER to treat these data
Equipements, plateaux techniques	<ul style="list-style-type: none"> • South Green : https://cat.opidor.fr/index.php/South Green • IDRIS : https://cat.opidor.fr/index.php/IDRIS

AgroLD Augmentation Dataset Benchmark	
4.1 Comment et avec quels moyens seront traitées les données ?	
Description	they will be treated manually

Anatomics	
4.1 Comment et avec quels moyens seront traitées les données ?	

IonoMil	
4.1 Comment et avec quels moyens seront traitées les données ?	
Description	we will develop computational scripts to treat datasets

AgronoMil	
4.1 Comment et avec quels moyens seront traitées les données ?	
Description	Python and R scripts will be developed to analyse the data
Equipements, plateaux techniques	<ul style="list-style-type: none"> • South Green : https://cat.opidor.fr/index.php/South Green

PlastiMil	
4.1 Comment et avec quels moyens seront traitées les données ?	
Description	Python and R scripts will be developed to analyse the data
Equipements, plateaux techniques	<ul style="list-style-type: none"> • South Green : https://cat.opidor.fr/index.php/South Green

Time Delay Correlation algorithm

4.1 Comment et avec quels moyens seront traitées les données ?

Description

We will use TDcor

Equipements, plateaux techniques

- South Green : <https://cat.opidor.fr/index.php/South Green>

PANTHEON

4.1 Comment et avec quels moyens seront traitées les données ?

Description

We will use the IRD I-TROP HPC

Equipements, plateaux techniques

- South Green : <https://cat.opidor.fr/index.php/South Green>

Lateral root dataset

4.1 Comment et avec quels moyens seront traitées les données ?

Description

Data will be analysed with TDcor and Pantheon and other methods developed in WP3

Equipements, plateaux techniques

- South Green : <https://cat.opidor.fr/index.php/South Green>

Time-series transcript profiling analysis in stem base of rice crown rootless 1 mutant after ectopic expression induction by dexamethasone of the CRL1 gene

4.1 Comment et avec quels moyens seront traitées les données ?

Description

we will use python scripts and HPC infrastructure

Equipements, plateaux techniques

- South Green : <https://cat.opidor.fr/index.php/South Green>

5. Stockage et sauvegarde des données pendant le processus de recherche

AgroLD Web application

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

Besoins de stockage AgroLD WebApp is a dabatase + a web application
all data will be stored at <http://agrold.org>

AgroLD_ETL

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

Besoins de stockage we will use GitHub repository

DLINKER

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

Besoins de stockage we wil use the GitHub repository <https://github.com/ANR-DIG-AI>

Biological Gated Recurrent Unit Transformer

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

Besoins de stockage we will use the GitHub repository <https://github.com/ANR-DIG-AI>

AgroLD_datasets

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

Besoins de stockage we will use the SouthGreen Bioinformatic and the zenodo
Volume estimé des données 5
Unité Go
Equipements, plateaux techniques

- South Green : [https://cat.opidor.fr/index.php/South Green](https://cat.opidor.fr/index.php/South%20Green)

OryzaGP

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

Besoins de stockage	Data will be stored on our dataverse repository https://dataverse.ird.fr/
Volume estimé des données	1
Unité	Go
Equipements, plateaux techniques	<ul style="list-style-type: none">• South Green : https://cat.opidor.fr/index.php/South Green

AgroLD Data Linking Benchmark

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

Besoins de stockage	Data will be stored on the project's Github
Volume estimé des données	200
Unité	Mo
Equipements, plateaux techniques	<ul style="list-style-type: none">• South Green : https://cat.opidor.fr/index.php/South Green

AgroLD Augmentation Dataset Benchmark

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

Besoins de stockage	they will be stored on the Github page
----------------------------	--

Anatomics

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

Besoins de stockage	datasets will be stored on our IT infrasctucture
Volume estimé des données	1
Unité	Go
Equipements, plateaux techniques	<ul style="list-style-type: none">• South Green : https://cat.opidor.fr/index.php/South Green

IonoMil

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

Besoins de stockage	Data will be stored on i-TROP DIADE storage cluster
Volume estimé des données	200
Unité	Go
Equipements, plateaux techniques	<ul style="list-style-type: none"> • South Green : https://cat.opidor.fr/index.php/South Green

AgronoMil

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

Besoins de stockage	Data will be store on the I-TROP HPC
Volume estimé des données	200
Unité	Go
Equipements, plateaux techniques	<ul style="list-style-type: none"> • South Green : https://cat.opidor.fr/index.php/South Green

PlastiMil

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

Besoins de stockage	Data will be stored on IRD I-Trop HPC
Volume estimé des données	200
Unité	Go
Equipements, plateaux techniques	<ul style="list-style-type: none"> • South Green : https://cat.opidor.fr/index.php/South Green

Time Delay Correlation algorithm

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

Besoins de stockage	TDcor is available on the CRAN archive and on IRD I-TROP cluster
Volume estimé des données	50
Unité	Mo
Equipements, plateaux techniques	<ul style="list-style-type: none"> • South Green : https://cat.opidor.fr/index.php/South Green

PANTHEON

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

Besoins de stockage	Data will be store on the IRD ITROP HPC
Volume estimé des données	200
Unité	Mo
Equipements, plateaux techniques	<ul style="list-style-type: none"> • South Green : https://cat.opidor.fr/index.php/South Green

Lateral root dataset	
5.1 Comment les données seront-elles stockées et sauvegardées tout au long du projet ?	
Besoins de stockage	Data will be stored on IRD Itrop HPC
Volume estimé des données	100
Unité	Mo
Equipements, plateaux techniques	<ul style="list-style-type: none"> • South Green : https://cat.opidor.fr/index.php/South Green

Time-series transcript profiling analysis in stem base of rice crown rootless 1 mutant after ectopic expression induction by dexamethasone of the CRL1 gene	
5.1 Comment les données seront-elles stockées et sauvegardées tout au long du projet ?	
Besoins de stockage	Datasets will be store in the core facilities
Volume estimé des données	100
Unité	Mo
Equipements, plateaux techniques	<ul style="list-style-type: none"> • South Green : https://cat.opidor.fr/index.php/South Green

6. Partage des données et conservation à long terme

AgroLD Web application	
6.1 Comment les données seront-elles partagées ?	
6.2 Comment les données seront-elles conservées à long terme ?	

Justification	Dumps of the datasets will be exported regularly and stored on Zenodo (see AgroLD dataset product)
----------------------	--

AgroLD_ETL	
6.1 Comment les données seront-elles partagées ?	
Modalités de partage	we will use GitHub repository
6.2 Comment les données seront-elles conservées à long terme ?	
Justification	we will use GitHub repository

DLINKER	
6.1 Comment les données seront-elles partagées ?	
Modalités de partage	Through the GitHub repository https://github.com/ANR-DIG-AI
6.2 Comment les données seront-elles conservées à long terme ?	
Justification	Archived in the GitHub repository https://github.com/ANR-DIG-AI

Biological Gated Recurrent Unit Transformer	
6.1 Comment les données seront-elles partagées ?	
Modalités de partage	we will share through the GitHub repository https://github.com/ANR-DIG-AI
6.2 Comment les données seront-elles conservées à long terme ?	
Justification	Will be archived on the GitHub repository

AgroLD_datasets	
6.1 Comment les données seront-elles partagées ?	

Modalités de partage

Datasets will be shared through the AgroLD Web application (www.agrold.org) and the zenodo platform <https://zenodo.org/record/4694518>

Entrepôt/Catalogue de données

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

6.2 Comment les données seront-elles conservées à long terme ?**Volume estimé des données**

5

Unité

Go

Date de début

2021-01-01

Archive

CINES Archivage numérique pérenne : https://cat.opidor.fr/index.php/CINES_Archivage_numérique_pérenne (CoreTrustSeal)

OryzaGP**6.1 Comment les données seront-elles partagées ?****Modalités de partage**

Datasets will be shared through the IRD dataverse portal <https://dataverse.ird.fr/>

Entrepôt/Catalogue de données

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

6.2 Comment les données seront-elles conservées à long terme ?**Justification**

non applicable

AgroLD Data Linking Benchmark**6.1 Comment les données seront-elles partagées ?****Modalités de partage**

data will be shared through our projet Github and through the IRD datasud repository

Entrepôt/Catalogue de données

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

6.2 Comment les données seront-elles conservées à long terme ?**Justification**

non applicable

AgroLD Augmentation Dataset Benchmark

6.1 Comment les données seront-elles partagées ?

Modalités de partage they will be shared on the Github and on the IRD Dataverse

6.2 Comment les données seront-elles conservées à long terme ?

Justification non applicable

Anatomics

6.1 Comment les données seront-elles partagées ?

Modalités de partage datasets will be shared through our dataverse

Entrepôt/Catalogue de données

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

6.2 Comment les données seront-elles conservées à long terme ?

IonoMil

6.1 Comment les données seront-elles partagées ?

Modalités de partage Data will be shared through the IRD dataverse

Entrepôt/Catalogue de données

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

6.2 Comment les données seront-elles conservées à long terme ?

Justification not defiend yet

AgronoMil

6.1 Comment les données seront-elles partagées ?

Modalités de partage

Data will be shared through the IRD dataverse

Entrepôt/Catalogue de données

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

6.2 Comment les données seront-elles conservées à long terme ?**PlastiMil****6.1 Comment les données seront-elles partagées ?****Modalités de partage**

Data will be shared through the IRD dataverse

Entrepôt/Catalogue de données

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

6.2 Comment les données seront-elles conservées à long terme ?

Question sans réponse.

Time Delay Correlation algorithm**6.1 Comment les données seront-elles partagées ?****Modalités de partage**

TDcor is available through the CRAN archive

6.2 Comment les données seront-elles conservées à long terme ?

Question sans réponse.

PANTHEON**6.1 Comment les données seront-elles partagées ?**

Modalités de partage

Pantheon will be shared through the GitHub pages

6.2 Comment les données seront-elles conservées à long terme ?

Question sans réponse.

Lateral root dataset**6.1 Comment les données seront-elles partagées ?****Modalités de partage**

Data will be shared through the IRD dataverse

Entrepôt/Catalogue de données

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

6.2 Comment les données seront-elles conservées à long terme ?

Question sans réponse.

Time-series transcript profiling analysis in stem base of rice crown rootless 1 mutant after ectopic expression induction by dexamethasone of the CRL1 gene**6.1 Comment les données seront-elles partagées ?****Modalités de partage**

we will use the IRD Dataverse

Entrepôt/Catalogue de données

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

6.2 Comment les données seront-elles conservées à long terme ?**Justification**

datasets will be hosted by NCBI