DMP du projet "BSWheat Marqueurs génétiques et caractères écophysiologiques pour la sélection de blés qui accumulent peu les contaminants métalliques"

Plan de gestion de données créé à l'aide de DMP OPIDoR, basé sur le modèle "ANR - Modèle de PGD (français)" fourni par Agence nationale de la recherche (ANR).

Renseignements sur le plan

Titre du plan	DMP du projet "BSWheat Marqueurs génétiques et caractères écophysiologiques pour la sélection de blés qui accumulent peu les contaminants métalliques"
Version	Version initiale
Domaines de recherche (selon classification de l'OCDE)	Agriculture, forestry, and fisheries
Langue	fra
Date de création	2023-03-07
Date de dernière modification	2023-06-21
Documents (publications, rapports, brevets, plan expérimental....), sites web associés	• Site internet du projet BSWheat : https://bswheat.hub.inrae.fr/ • Compte-rendu de l'assemblée générale de lancement : Diapo_AG1_BSWheat_et_compte_rendu.pdf • Compte-rendu visioconférence point étape WP1.1 16 mai 2023 : Point_etape_16_mai_2023_avec CR.pdf

Renseignements sur le projet

Titre du projet	BSWheat Marqueurs génétiques et caractères écophysiologiques pour la sélection de blés qui accumulent peu les contaminants métalliques
Acronyme	BSWheat
Résumé	Le projet B-SWheat vise à identifier des marqueurs génétiques et des caractères écophysiologiques chez le blé dur et le blé tendre afin d’être utilisés par les sélectionneurs pour produire des cultivars accumulant peu d’éléments traces (ET) toxiques et pour certains réglementés (As, Cd, Pb, Ni), sans impacter le rendement, la teneur des grains en oligoéléments (Fe, Zn) et la bio-accessibilité digestive de ces derniers. Le projet combine des études d'associations à l'échelle du génome, le criblage de la diversité allélique de gènes ciblés et le phénotypage au champs de caractères écophysiologiques liés aux système racinaire, à la

répartition de la biomasse, à la proportion des tissus de transport/stockage et au fonctionnement de la plante, notamment la mobilisation et l'absorption des ET dans la rhizosphère, la phénologie, la remobilisation des nutriments. Une attention particulière sera portée à l'allocation des ET à et au sein de l'épi en lien avec le rendement. Les marqueurs génétiques et les caractères écophysiologiques corrélés à la teneur en ET du grain seront testés aux champs 2 ans de suite sur 4 sites expérimentaux pour évaluer la stabilité de leur liaisons avec les ET du grain dans des contextes pédoclimatiques et de disponibilité des ET du sol variés. L'importance des effets des marqueurs et caractères retenus sur la qualité sanitaire et nutritive du grain sera quantifiée ce qui permettra ainsi de les hiérarchiser en prenant en compte également la stabilité de leur effets selon l'environnement. Grâce à l'expertise des deux sélectionneurs de blé privés du projet et d'Arvalis, nous déterminerons finalement quels marqueurs génétiques et écophysiologiques peuvent être intégrés dans les schémas de sélection variétale. La caractérisation des variétés de blés actuellement cultivées selon ces marqueurs sera diffusée auprès des agriculteurs pour les aider rapidement à produire du blé moins contaminés. Une synthèse du projet permettra d'évaluer le gain possible en matière de sécurité sanitaire du blé et les compromis associés sur le rendement ou la valeur nutritionnelle. Ces informations seront communiquées aux acteurs de la production de blé, aux autorités sanitaires et réglementaires avec qui les partenaires du projet travaillent.

Sources de financement

- Agence Nationale de la Recherche : ANR-22-CE34-0023

Date de début 2023-01-01

Date de fin 2027-12-31

Partenaires

- ARVALIS INSTITUT DU VEGETAL
- SOCIETE FLORIMOND DESPREZ VEUVE ET FILS
- Université Pau et Pays de l'Adour
- RAGT 2N
- Unité de Service Analyses Végétales et Environnementales
- Génétique Diversité et Ecophysiologie des Céréales
- Interaction Sol Plante Atmosphère

Produits de recherche :

1. Fichiers bruts des analyses d'éléments traces et majeurs (Jeu de données)
2. Code de traitement des données au format R (Logiciel)
3. Caractérisation génétique de génotypes de blés selon la puce SNP 420K (Jeu de données)
4. Fichiers de données prétraitées prêtes à être analysées (Jeu de données)
5. Caractérisation génétique de génotypes de blé dur (puce SNP 420K) et identification / développement de marqueurs associés (Jeu de données)
6. Test de quelques génotypes (blé tendre et dur) dans des conditions environnementales contrastées (Jeu de données)

Contributeurs

Nom	Affiliation	Rôles
DEMENOU Boris - https://orcid.org/0000-0002-3308-1391	ARVALIS Institut du Végétal, Service SAGEP	• Personne contact pour les données (Génotypage et GWAS, Test markers & trait)
Le Gouis Jacques - https://orcid.org/0000-0001-5726-4902	INRAE Génétique Diversité et Ecophysiologie des Céréales - 200017467R	• Personne contact pour les données (Génotypage)
NGUYEN Christophe - http://orcid.org/0000-0003-0917-6398	Interaction Sol Plante Atmosphère - 201421781U	• Coordinateur du projet • Personne contact pour les données (Code informatique R, Données pour analyse) • Responsable du plan de gestion de données
Soulé Patrice	Inrae Unité de Service Analyses Végétales et Environnementales - 200122726C	• Personne contact pour les données (Analyse élémentaire)

Droits d'auteur :

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

DMP du projet "BSWheat Marqueurs génétiques et caractères écophysiologiques pour la sélection de blés qui accumulent peu les contaminants métalliques"

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

1a. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?

Les nouvelles données seront produites de la manière suivantes:

- Expérimentation sur végétal au champ ou en conditions contrôlées visant à comparer des variétés de blé ou à moduler leur écophysiologie pour identifier les conséquences sur leur teneur en éléments traces métalliques (ETM)
- Analyses du matériel végétal (biomasses, caractères morphologiques, stades de développement, composition chimique)
- Analyses génétiques: recherche de séquences spécifiques dans le génome de variétés de blé dur et tendre
- Modélisation: statistiques ou mécanistes permettant d'identifier des relations entre variables caractérisant le végétal et la teneur en éléments traces des grains

Les nouvelles données sont issues de travaux répondant à des protocoles expérimentaux. Elles feront l'objet de rapports écrits relatifs à leur production pour présentation aux partenaires et *in fine* leur publication sous forme d'articles scientifiques explicitera leur provenance.

Données préexistantes:

- Collections de génotypes de variétés de blé: les variétés seront comparées pour étudier leur variabilité d'accumulation d'ETM
- Données de génotypage (SNP) et de quelques caractéristiques phénotypiques: elles seront utilisées pour rechercher des corrélations avec la variabilité des teneurs en ETM des variétés.

Ces données ont fait l'objet de publications décrivant leur provenance et les méthodes d'obtention

---

1b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?

- Données numériques issues d'analyses chimiques, physiques, de mesures de biomasses sur des échantillons végétaux et de terre: fichiers tableur (Excel, libre office calc) et fichier texte CSV, séparateur virgule
- Rapports: format pdf, doc, docx, odt: rapport d'analyses, compte-rendus de réunions, protocoles
- Publications: format pdf, doc, docx, odt

---

2. Documentation et qualité des données

2a. Quelles métadonnées et quelle documentation (par exemple méthodologie de collecte et mode d'organisation des données) accompagneront les données ?

Les données du projet sont déposées sur l'espace partagé https://nextcloud.inrae.fr. Ce dernier contient un dossier par work package (y compris le WP0 de coordination) et au sein de chaque dossier de work package, les dossiers sont organisés par action et sont nommés de manière explicite pour une identification facile

Chaque fichier partagé est accompagné d'un fichier texte du même nom avec le suffixe "lisezmoi" qui contient les métadonnées selon le format Dublin Core: Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage, Rights

Les fichiers de résultats sont accompagnés des fichiers de protocoles expérimentaux et/ou analytiques

Seuls les fichiers finaux sont déposés sur le site. Les versions de travail ne sont pas déposées

2b. Quelles mesures de contrôle de la qualité des données seront mises en œuvre ?

Tout fichier déposé sur l'espace partagé fait l'objet d'une double vérification: par la personne produisant le fichier et par le coordinateur qui vérifie le dépôt.

Pour les données numériques, on recherche les erreurs 'visibles directement' (cohérence du fichier, homogénéité des séparateurs décimaux, erreur de typographie etc.) et une analyse statistique examine les valeurs pour identifier des valeurs suspectes qui demande une vérification

Le coordinateur vérifie également que chaque fichier est bien accompagné de son protocole d'obtention et des fichiers 'lisezmoi' associés.

Pour les analyses chimiques, elles sont effectuées par le partenaire Usrave qui est accrédité Cofrac. Pour les quelques analyses qui ne seront pas réalisées par l'Usrave, les incertitudes, limites de quantification sont calculées et chaque série d'analyses comprend des échantillons de références.

3. Stockage et sauvegarde pendant le processus de recherche

3a. Comment les données et les métadonnées seront-elles stockées et sauvegardées tout au long du processus de recherche ?

Les données du projets sont déposées sur un cloud 'Nextcloud' sauvegardé par Inrae. En outre, la copie est également stockée sur l'ordinateur du coordinateur qui est lui-même doublement sauvegardé: par l'unité Inrae Ispa sur ses serveurs et par le coordinateur (sauvegarde journalière sur disque dur externe).

3b. Comment la sécurité des données et la protection des données sensibles seront-elles assurées tout au long du processus de recherche ?

Il n'y a pas de données véritablement sensibles. Les données déposées sur l'espace partagé Nextcloud sont sécurisées à la fois par les protocoles du service Nextcloud et par ceux d'Inrae.

Les données stockées sur les ordinateurs personnels sont protégées par un code de verrouillage de l'ordinateur

4. Exigences légales et éthiques, codes de conduite

4a. Si des données à caractère personnel sont traitées, comment le respect des dispositions de la législation sur les données à caractère personnel et sur la sécurité des données sera-t-il assuré ?

Cette question est traitée par un accord de consortium qui a été établi entre les partenaires.

Il n'y a pas de données véritablement personnelles dans le projet. Des essais agronomiques seront conduits chez des agriculteurs avec leur accord mais les publications techniques et scientifiques ne feront pas référence à des données personnelles.

Les participants du projet se voient cités sur le site internet du projet mais les données qui y figurent sont strictement professionnelles et sont publiées avec l'accord des intéressés

4b. 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 ?

Cette question est traitée par l'accord de consortium qui a été établi entre les partenaires.

4c. Comment les éventuelles questions éthiques seront-elles prises en compte, les codes déontologiques respectés ?

Cette question est traitée par un accord de consortium qui a été établi entre les partenaires.

---

5. Partage des données et conservation à long terme

5a. Comment et quand les données seront-elles partagées ? Y-a-t-il des restrictions au partage des données ou des raisons de définir un embargo ?

Cette question est traitée par l'accord de consortium qui a été établi entre les partenaires.

5b. Comment les données à conserver seront-elles sélectionnées et où seront-elles préservées sur le long terme (par ex. un entrepôt de données ou une archive) ?

Une fois les données utilisées et publiées sous forme d'articles de recherche scientifiques, les données seront déposées sur l'entrepôt: https://entrepot.recherche.data.gouv.fr/dataverse/inrae avec les métadonnées associées

---

5c. Quelles méthodes ou quels outils logiciels seront nécessaires pour accéder et utiliser les données ?

Les données déposées sur l'entrepôt https://entrepot.recherche.data.gouv.fr/dataverse/inrae seront en accès libre (open data) avec les métadonnées les décrivant. Le format texte avec séparateur virgule (format csv) sera utilisés pour les données numériques.

Si des données textuelles doivent être déposées, elles le seront également sous le format basique binaire texte (.txt) au format non propriétaire.

5d. Comment l'attribution d'un identifiant unique et pérenne (comme le DOI) sera-t-elle assurée pour chaque jeu de données ?

L'identifiant pérenne (DOI) sera attribué automatiquement lors de la publication scientifique des articles par les revues et par l’entrepôt de dépôt des données https://entrepot.recherche.data.gouv.fr/dataverse/inrae

---

6. Responsabilités et ressources en matière de gestion des données

6a. Qui (par exemple rôle, position et institution de rattachement) sera responsable de la gestion des données (c'est-à-dire le gestionnaire des données) ?

Chaque responsable partenaire est responsable de la gestion des données produites par sa structure : vérification du format, production des métadonnées associées et dépôt sur le site partagé du projet et ensuite sur l'entrepôt https://entrepot.recherche.data.gouv.fr/dataverse/inrae.

Ces personnes sont identifiées comme contributeur au PGD.

Le coordinateur s'assurera de la conformité de la démarche avec le plan de gestion des données qui sera progressivement complété au cours du projet

---

6b. Quelles seront les ressources (budget et temps alloués) dédiées à la gestion des données permettant de s'assurer que les données seront FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable) ?

Les outils de partage et de stockage des données sont des services publics gratuits et il n'y a donc pas de budget du projet à y consacrer. Le coordinateur dégagera de son temps pour que chaque producteur de données accomplisse les tâches de gestion des données identifiées dans le plan de gestion des données