
DMP du projet "DEFI : Données pour l'Education, la Formation, l'Innovation" | Version initiale

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 "DEFI : Données pour l'Education, la Formation, l'Innovation" Version initiale
Version	Version initiale
Objet/périmètre du plan	<p>Ce document de cadrage est réalisé dans le cadre de l'organisation d'un groupe thématique numérique (GTnum) de la Direction du numérique pour l'éducation (Ministère de l'éducation nationale, de la jeunesse et des sports) qui prévoit la réalisation d'un DMP avec l'outil DMP OPIDOR pour tous recueils et traitements de données.</p> <p>C'est aussi une première phase de construction d'un data paper.</p>
Domaines de recherche (selon classification de l'OCDE)	Educational sciences, Computer and information sciences
Langue	fra
Date de création	2022-07-26
Date de dernière modification	2022-07-27
Documents (publications, rapports, brevets, plan expérimental....), sites web associés	<ul style="list-style-type: none">Bibliographie collaborative du projet DEFI : https://www.zotero.org/groups/4574772/gtnumdefi/library

Renseignements sur le projet

Titre du projet	Données pour l'Education, la Formation, l'Innovation
Acronyme	DEFI

Résumé

Le projet s'attache au développement de la littératie des données chez les enseignants et les élèves. Cette littératie des données intègre la capacité à comprendre les enjeux de la production, de l'organisation et de l'exploitation des données, et à les utiliser efficacement et de manière critique et créative dans le contexte scolaire. Elle fait partie des compétences nécessaires pour évaluer et utiliser l'information, intégrées dans le Cadre de référence des compétences numériques.

Le projet DEFI repose sur une recherche-intervention visant à transformer les pratiques individuelles et organisationnelles des acteurs, en prenant appui sur des travaux de recherche de plusieurs équipes, qui permettront de comprendre les représentations, et d'évaluer l'intérêt du recours aux données pour enseigner et éduquer.

Plus d'infos : <https://edunumrech.hypotheses.org/4185>

Sources de financement

- Ministère de l'éducation nationale, de la jeunesse et des sports – Direction du numérique pour l'éducation :

Date de début

2022-01-01

Date de fin

2025-01-01

Partenaires

- INSPE BORDEAUX ()
- IMS, UMR 5218 CNRS - Université de Bordeaux (200711887V)
- DANE Bordeaux ()
- DRANE de la Région Nouvelle Aquitaine ()
- MICA-ICIN, Université Bordeaux Montaigne (200919217D)
- PASSAGES, UMR 5319 CNRS - Université Bordeaux Montaigne (201622032H)
- LABPSY, EA 4139 - Université de Bordeaux (200715383V)
- BSE, UMR 6060 CNRS - Université de Bordeaux (202224195H)
- CEMTI, EA 3388 - Université Paris 8 (200114665R)
- LIFAT, EA 6300 - Université de Tours (201220254T)
- Atelier Canopé 35 ()
- Atelier Canopé 94 ()
- Atelier Canopé 13 ()
- CLEMI (Centre pour l'éducation aux médias et à l'information) ()
- Réseau Canopé, Direction de la Recherche et du Développement ()
- TECHNE, EA 6316 - Université de Poitiers (201220297P)
- Groupement d'Intérêt Scientifique 2iF ()
- ACté, EA 4281 Université Clermont-Auvergne (200815518M)
- INSPE CLERMONT AUVERGNE ()
- LERASS, EA 827 - Université Paul Valéry - Université Montpellier 3 (199113245R)
- INSPE CRETEIL ()
- CEDITEC, EA 3119 - Université Paris Est (199914429T)

Produits de recherche :

1. Default research output

Contributeurs

Nom	Affiliation	Rôles
CAPELLE Camille - idref 16855707X	IMS, UMR 5218 CNRS - Université de Bordeaux	<ul style="list-style-type: none">• Coordinateur du projet• Personne contact pour les données
ONFROY Karine - https://orcid.org/0000-0003-2081-029X	BSE, UMR 6060 CNRS - Université de Bordeaux	<ul style="list-style-type: none">• Responsable du plan de gestion de données

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 "DEFI : Données pour l'Education, la Formation, l'Innovation" | Version initiale

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

Le projet proposé produira plusieurs types de données dans divers formats mais aussi une documentation facilitant leur réutilisation et des ressources pédagogiques à destination des enseignants et des élèves.

L'utilisation de formats ouverts sera privilégiée pour que les fichiers restent indépendants d'une technologie unique, donc lisibles et réutilisables dans le temps.

L'ensemble sera stocké sur des espaces sécurisés durant tout le processus de recherche (Espace Tribu, Cloud de l'IMS, Serveurs de l'Université de Bordeaux et de la TGIIR Huma_Num).

IDENTIFICATION DES JEUX DE DONNEES (quanti. et quali)

- **Les données issues de l'enquête en ligne**

Près de 3000 répondants sont attendus. Ils sont contactés grâce à différents supports pour multiplier les profils : Réseaux sociaux et mailing liste constituée lors des formations. Les réponses sont collectées via le logiciel Sphinx Online et hébergées sur les serveurs de l'Université. La base de données obtenue sera enregistrée au format ouvert CSV et se présentera sous la forme d'un tableau avec 1 ligne par répondant et chacune de leurs réponses en colonnes.

- **Les données des entretiens**

30 entretiens d'une durée approximative d'1h vont être menés auprès de volontaires recrutés par le biais du questionnaire en ligne mais aussi à l'issue des formations. Les interviews seront enregistrés (audio) et transcrits en texte. Les données personnelles recueillies seront pseudo-anonymisées avant d'être analysées. La table de correspondance permettant d'identifier les individus sera stockée dans l'espace personnel de la responsable de projet (Cloud du laboratoire IMS). Il en est de même pour les formulaires de consentement dans le cadre de la collecte de données personnelles signés par les interviewés. Les fichiers de transcriptions et la grille d'entretien seront enregistrés au format ouvert TXT. Le codage et l'analyse des entretiens se feront avec le logiciel MAXQDA.

- **Les données d'observation**

Les informations recueillies par le biais des méthodes d'observation (essentiellement des schémas de l'organisation de l'espace réalisés à main levée) seront numérisés au format ouvert JPG.

- **Les photos**

Près de 1000 photos seront prises durant les séances d'observation. Pour ce corpus iconographique, le format ouvert JPG sera privilégié pour les fichiers. Un logiciel comme TROPY facilitera la gestion des images et leur description.

- **Les productions des étudiants/élèves**

Durant les formations, les enquêtés vont produire des Dataviz ou d'autres types de productions qui vont être sauvegardées sous forme de fichiers image (format ouvert).

- **Les carnets de bord des travaux de groupes d'étudiants/d'élèves**

La démarche qui a permis de produire les productions des étudiants/élèves est recueillie dans un carnet de bord en ligne (nootebook au format numérique). Les informations seront enregistrées dans un tableur au format ouvert CSV.

- **Les scénarios de formation collectés sur le web.**

Des scénarios de formation en lien avec le numérique seront collectés sur le web (site EDU'Base), par des techniques de webscraping. Les formats ouverts seront privilégiés pour les fichiers de sortie.

- **Les scénarios de formation créés durant le projet**

Les membres du projet et les enseignants prévoient de créer une trentaine de scénarios (fichiers texte de 2-3 pages chacun).

- **La bibliographie collaborative du projet**

L'état de l'art des recherches sur la littératie des données en éducation et le travail de veille feront l'objet d'une bibliographie dédiée via l'outil Zotero. En accès ouvert dès le début du projet, la bibliothèque en ligne va faciliter le travail collaboratif et les ressourcements scientifiques.

Bibliographie collaborative et ouverte du projet DEFI : <https://www.zotero.org/groups/4574772/gtnumdefi/library>

2. Documentation et qualité des données

Les informations nécessaires à la compréhension et réutilisation de chacun des jeux de données listés ci-dessus seront enregistrées dans des **fichiers texte de type « lisez-moi » et dans des dictionnaires de variables et modalités (codebook)**. Ils détailleront la méthodologie utilisée pour collecter les données, pour les analyser ...

Une convention de nommage des fichiers a été mise en place dès le début du projet par type de données (scénarios de formation, notes d'observation, carnets de bord des participants ...).

Un lexique (livrable) constitué à partir des entrées thématiques du projet de recherche est enrichi tout au long du projet de manière collaborative.

La responsable du projet sera chargée de contrôler la qualité des données au niveau de leur description mais aussi au niveau de la donnée en elle-même (valeurs nulles, doublons, valeurs anormales ..) grâce à des validations décidées avec les responsables de chaque axe lors de réunions dédiées. Des outils comme FACILE du CINES permettront de contrôler la qualité des données. La mise en place de cette **démarche organisationnelle** est en réflexion.

3. Stockage et sauvegarde pendant le processus de recherche

Les modalités de stockage et de sauvegarde envisagées s'appuient sur des outils hébergeant les données en France :

- l'écosystème Huma_Num (TGIR pour les SHS)
- l'espace collaboratif de l'Education nationale TRIBU
- le cloud de l'IMS
- les serveurs de l'Université (SphinxOnline)

L'accès à ces différents espaces est contrôlé par la responsable du projet.

QUELLES SOLUTIONS POUR QUELS TYPES DE DONNEES ?

- **1. ShareDoc (outil proposé par Huma_Num)**

pour le stockage et le partage des données chaudes entre tous les membres du projet

--> sets de données, documentations sur les données

- **2. TRibu (plateforme de partage de documents proposée par le Ministère de l'Education Nationale)**

pour le partage des documents du projet entre tous les membres

--> documents de cadrage, documents administratifs, livrables

- **3. Cloud de l'IMS (espace personnel de la responsable du projet)**

pour sauvegarder les données (1) et les documents du projet (2) dans un second espace sécurisé mais aussi les documents relatifs à la protection des données et au respect de la vie privée

--> sets de données, documentations sur les données

--> documents de cadrage, documents administratifs, livrables

--> formulaires signés par les participants au projet (consentements et droits à l'image) et la table de correspondance entre les données directement identifiantes et leur alias (numéro séquentiel ..)

- **4. Serveurs de l'Université de Bordeaux**

La solution SphinxOnline est proposée par l'Université pour la réalisation d'enquêtes en ligne. Durant la passation, les données sont hébergées sur les serveurs de l'institution.

--> réponses de l'enquête en ligne durant son ouverture

4. Exigences légales et éthiques, codes de conduite

De nombreuses données à caractères personnel sont traitées au sein du projet. La base légale du traitement repose sur l'exécution d'une mission de recherche publique.

Le Délégué à la Protection des Données personnelles de l'Université de Bordeaux a été sollicité et les membres du projet s'engagent à respecter la législation en vigueur tout au long du projet DEF1 (RGPD, droit à l'image ...).

5. Partage des données et conservation à long terme

Toute les données produites seront partagées librement afin de promouvoir la science ouverte (sets de données, documentation associée, travail de veille ...).

On privilégiera l'utilisation d'outils de diffusion d'entrepôts de diffusion reconnus comme NAKALA (outil Huma_Num). Le projet utilisera au maximum des formats ouverts utilisables avec des applications libres, aucune donnée finale ne nécessitera de logiciel particulier pour son utilisation. Toutes les données diffusées à l'issue du projet seront déposées sur des entrepôts ou archives fonctionnant par création d'un identifiant unique et pérenne.

Notes

ZOTERO : logiciel de gestion de références gratuit, libre et open source qui s'inscrit dans la philosophie du Web 2.0. Il permet de gérer des données bibliographiques et des documents de recherche. Il a été développé par le "Center for History and New Media" de l'université George Mason (USA). Plus d'infos ici : <https://zotero.hypotheses.org/>

TROPY : application open source qui aide les chercheurs à décrire et gérer leurs photos. Elle a été développée par le "Center for History and New Media" de l'université George Mason à qui l'on doit déjà ZOTERO. Plus d'infos ici : <https://docs.tropy.org/#tropy>

TRIBU : Tribu est un service qui permet à un groupe d'utilisateurs de partager un espace collaboratif sécurisé dédié à un projet. Tribu permet à la communauté de partager des documents, des agendas, des tâches, des forums de discussion. Il offre la possibilité d'être libre de son organisation et permet de travailler en ligne. Plus d'infos ici : <https://tribu.phm.education.gouv.fr/>

HUMA_NUM : la Très Grande Infrastructure de Recherche (TGIR) des Humanités numériques. Plus d'infos ici : <https://www.huma-num.fr/presentation/>

SHAREDOCS : gestionnaire de fichiers mis en oeuvre par la TGIR Huma-Num, sur ses propres serveurs. Il est adapté aux projets de recherche qui souhaitent stocker, échanger, partager, travailler sur des données de type fichiers (ensemble de photos, de textes transcrits, etc). Ainsi, l'usage courant de ShareDocs permet de préparer des fichiers pour une édition en ligne ou bien une diffusion dans **NAKALA** (par exemple). Plus d'infos ici : <https://documentation.huma-num.fr/sharedocs-stockage/>

NAKALA : Nakala est un service d'Huma-Num permettant à des chercheurs, enseignants-chercheurs et équipes de recherche de partager, publier et valoriser tous types de données numériques documentées (fichiers textes, sons, images, vidéos, objets 3D, etc.) dans un entrepôt sécurisé afin de les publier en accord avec les principes de la *FAIR data* (Facile à trouver, Accessible, Interopérable et Réutilisable). Plus d'infos ici : <https://documentation.huma-num.fr/nakala/>

6. Responsabilités et ressources en matière de gestion des données

Tout au long du projet, chaque responsable d'axe gère les données qui lui sont associé :

Axe 1: Anne Lehmans (IMS, UMR 5218 CNRS - Université de Bordeaux)

Axe 2 : Camille Capelle (IMS, UMR 5218 CNRS - Université de Bordeaux)

Axe 3 : Béatrice Drot-Delange (ACté, EA 4281 Université Clermont-Auvergne)
