



HAL
open science

DMP du projet “ Plan de gestion de données du GT NUM ELN ”

Anne Cordier, Laura Pauchard

► **To cite this version:**

Anne Cordier, Laura Pauchard. DMP du projet “ Plan de gestion de données du GT NUM ELN ”. Université de Lorraine. 2023. hal-04318577

HAL Id: hal-04318577

<https://hal.univ-lorraine.fr/hal-04318577v1>

Submitted on 5 Dec 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.



Open licence - etalab

DMP du projet "Plan de gestion de données du GT NUM ELN"

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 "Plan de gestion de données du GT NUM ELN"
Langue	fra
Date de création	2021-03-11
Date de dernière modification	2021-03-17
Identifiant	PGD_GTNUM_ELN_2020-2022

Renseignements sur le projet

Titre du projet Plan de gestion de données du GT NUM ELN

Résumé Le projet vise à répondre aux questions suivantes :
Comment accompagner la transformation numérique ainsi que la reconfiguration des espaces d'apprentissage (école / domicile) dans un contexte socialement et territorialement différencié/hétérogène (identifier et repérer les disparités, développer de nouveaux leviers d'action pour y remédier).

Sources de financement

- Direction du Numérique pour l'Education :

Produits de recherche :

1. Default research output (Jeu de données)

Contributeurs

Nom	Affiliation	Rôles
Anne Cordier		<ul style="list-style-type: none">• Coordinateur du projet• Personne contact pour les données
Laura Pauchard		<ul style="list-style-type: none">• Responsable du plan

DMP du projet "Plan de gestion de données du GT NUM ELN"

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 ?

Plusieurs nouvelles données seront produites dans le cadre de ce programme de recherche.

D'abord, un travail d'enquête par questionnaire sera réalisé auprès de parents d'élèves afin de réaliser un premier diagnostic sur le niveau d'équipement numérique des familles et les pratiques au sein du foyer. Ce questionnaire sera établi en ligne et circulera à l'échelle nationale, via différents réseaux : via la DNE et la DANE, les établissements scolaires, les associations de parents-d'élèves, les réseaux professionnels des chercheurs du programme. Les réponses au questionnaire seront gérées et traitées via Qualtrix survey (Université de Toulouse).

Ensuite, plusieurs établissements scolaires seront choisis, au regard des résultats de cette enquête et au sein de cinq régions d'études pressenties du fait des contacts professionnels des chercheurs (Normandie, Nouvelle Aquitaine, Occitanie, Bretagne, Hauts-de-France), pour une étude qualitative. Cette phase s'appuiera sur une phase d'entretiens auprès d'élèves, de parents et d'enseignants dans une quinzaine d'établissements. Il visera à documenter les pratiques et les représentations du numérique dans la classe et hors la classe. Elle s'appuiera sur la réalisation de monographies qui seront composées de plusieurs temps :

Sous-tâche 1 : Entretien avec la Direction de l'École, voire l'équipe pédagogique de l'école

Sous-tâche 2 : Entretien avec l'Enseignant-e de la classe ciblée

Sous-tâche 3 : Observation distanciée de séances pédagogiques engageant le numérique

Sous-tâche 4 : Séance (Observation participante) en recherche-intervention sur "Toi et le numérique"

Sous-tâche 5 : Focus Group avec les Parents d'élèves scolarisés en cycles 2 et 3

Dans le cadre de cette seconde phase de la recherche, les données collectées seront notamment la production de dessins/collage d'enfants et de verbatims autour de ces derniers pour analyser leurs représentations et leur écosystème numérique personnel.

Les pratiques numériques enfantines ont déjà été étudiées par les anglosaxons, mais ce programme appréhende ces questions sous l'angle de la confrontation entre école/famille/territoires.

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

En ce qui concerne le questionnaire, pour chaque réponse nous disposerons du nombre d'enfants, de leur sexe, de leur âge, de leur niveau scolaire, de la catégorie socio-professionnelle des parents, et du code postal du foyer. A cela s'ajouteront des informations sur les équipements numériques, le niveau de connexion et les pratiques. La période de recueil des réponses au questionnaire sera d'une durée de deux mois. On ignore aujourd'hui le nombre de réponses possibles à cette enquête.

Pour la phase qualitative, les données produites seront des fichiers audios et vidéos des entretiens et focus group ainsi que des fichiers texte des entretiens transcrits. Les observations recueillies dans un carnet de terrain seront retranscrites dans des fichiers textes. Enfin, des dessins-collage d'enfants pourront être collectés et numérisés (jpeg) dans le cadre de la phase d'expérimentation. Des photographies des expérimentations en classe seront réalisées (jpeg). Une quinzaine d'établissements scolaires seront étudiés.

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 ?

Le standard de métadonnées le plus adapté semble la DDI pour ce qui relève des enquêtes quantitatives et qualitatives. Le Dublin Core sera peut-être plus opportun pour certains type de données produites.

Des documents seront rédigés pour documenter les métadonnées des enquêtes et le protocole méthodologique (que ce soit pour la phase quantitative ou qualitative). Des notices d'informations documenteront la méthodologie de la collecte de données, les méthodes d'analyses, etc.

En ce qui concerne l'enquête par questionnaire, nous pourrions renseigner : la couverture géographique, le nombre de réponses, les variables, la période de l'enquête, support de l'enquête (enquête en ligne) et condition de diffusion (comment le questionnaire a été diffusé dans les réseaux), un résumé de l'enquête et ses objectifs. L'élaboration d'un tel fichier de métadonnées pourra être aussi envisagé sur la production de dessins/collage des enfants.

Une convention de nommage des fichiers sera prochainement définie avec l'équipe pour structurer l'espace partagé de la Box Huma-num et s'y

retrouver dans l'organisation des fichiers.

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

Chaque protocole méthodologique est discuté collectivement par les membres scientifiques du GT Num. Un protocole commun est en cours de définition pour les deux phases de la recherche (quantitative et qualitative). Pour la phase d'étude sur le terrain, chaque membre scientifique, en fonction du contexte du terrain étudié, pourra ajuster le protocole et en faire part aux autres membres.

Les terrains d'études seront ciblés de deux manières : d'une part, en fonction de critères socioéconomiques et territoriaux propres à la problématique scientifique, d'autre part en fonction des résultats de l'enquête par questionnaire, sur cinq régions déjà étudiées par les chercheurs et avec qui ils et elles ont déjà des contacts avec les établissements.

Par ailleurs, les autorisations des tutelles institutionnelles de référence (DANE pour le second degré et DASEN pour le premier degré) ont été obtenues pour chaque territoire d'étude.

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 seront stockées et sauvegardées sur la BOX Huma-Num de Nantes via l'accompagnement de Stéphane Loret, ingénieur à l'UMR ESO. Les membres scientifiques du programme de recherche auront accès aux données stockées sur la BOX. Une authentification leur sera demandée et leur accès sera validée par la porteuse du programme de recherche, Anne Cordier.

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

La récupération des données est assurée par la TGIR Huma-Num en cas d'incident.

L'accès à la Box Huma Num sera géré par la responsable scientifique Anne Cordier. L'accès nécessite une authentification par les membres du programme. Les données seront accessibles uniquement aux membres scientifiques du projet de recherche.

Au regard de la sensibilité des données, un travail d'anonymisation ou de pseudonymisation pourra être envisagé afin de réduire les risques. Si les catégories de personnes sont sensibles au regard de la loi, l'impact sur la vie privée de ces enfants reste limité au regard du sujet d'étude (les pratiques, représentations et équipements numériques).

La protection des données sera assurée par un système d'accès sécurisé aux données et une instruction du dossier pour la conformité RGPD nous permettra de renforcer la sécurité si celle-ci est jugée insuffisante.

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 ?

L'instruction pour la mise en conformité RGPD du programme GTNUM est en cours. Le dossier sera instruit par Emilie Masson (DPO CNRS). En effet, le programme va recueillir des données indirectement identifiantes pour ce qui relève de l'enquête par questionnaire et identifiantes directement pour ce qui relève des travaux de recueil de données qualitatives (entretiens et observations).

Le code postal est demandé dans l'enquête par questionnaire, et dans laquelle des données sur l'âge, le nombre d'enfants et la CSP des parents seront aussi collectés. Il y a donc des données à caractère personnel indirectement identifiantes, dans ce premier temps de production de données. Un encart RGPD précisant les droits des enquêtés ainsi que les personnes à contacter pour exercer ces droits sera rédigé au début du

questionnaire en ligne. De la même manière, les répondants au questionnaire seront bien informés des objectifs de la recherche afin d'obtenir un consentement éclairé.

Une attention particulière est portée à l'égard des catégories d'enquêtés puisque pour une partie d'entre eux, il s'agit d'enfants de 6 à 11 ans. Les chercheurs, habitués de ce public et des questions éthiques et juridiques de par leurs expériences scientifiques en milieux scolaires, mettront tout en œuvre pour s'assurer du respect de la réglementation et des recommandations à venir de la DPO.

Un travail de floutage des visages d'enfants sera réalisé sur les photographies. Le consentement à la participation vidéo dans les focus group sera distingué du consentement sur la production de données de recherche.

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 ?

Au sein de l'équipe de recherche, une charte d'utilisation des données du GT Num a été discutée et est en cours de rédaction afin de préciser les conditions d'utilisations des données produites par chacun des membres. L'objectif étant de partager l'ensemble des données produites tout en encadrant leur utilisation (citation des membres scientifiques ayant concouru à la production de la donnée, association des membres dans l'écriture des articles, etc.).

Une réflexion sera engagée au sein de l'équipe afin d'appliquer des licences creatives commons sur les jeux de données produits.

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

Les membres scientifiques du programme mènent tous des recherches impliquant la personne humaine, et de surcroit des mineurs. Ils connaissent et appliquent la loi Jardé (décret du 17/11/2016). Pour ce qui concerne l'accès au public-cible, les chercheurs de l'équipe sont habitués à enquêter dans des classes, et pour ce faire ont déjà développé des relations de confiance avec les Rectorats et notamment les DANÉ, ce qui constitue un facteur favorable à l'accès aux terrains, en sus de l'habitude à investiguer auprès d'un public d'élèves et d'enseignants.

Parce qu'une partie des données concernent des mineurs, la vigilance en matière de stockage et de sécurisation de l'accès aux données se trouve renforcée.

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 ?

Une fois les données produites, analysées et valorisées (horizon 2024), l'équipe reste ouverte au partage des données pourvu qu'elle en est les moyens humains et techniques. Elle projette d'organiser des temps d'échanges à ce sujet, au cours du programme.

Une partie de l'équipe souhaite se former quant à l'archivage des données (fonctions, méthodes, outils, réglementations) pour envisager plus en détails ces possibilités.

Les éventuelles restrictions en matière de partage des données tiennent à la possibilité ou non d'anonymiser les données.

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) ?

Les données produites peuvent présenter un intérêt scientifique pour tous les chercheurs impliqués dans les recherches sur l'éducation, la parentalité, et les pratiques numériques. Les données seront conservées et partagées selon la possibilité ou non d'anonymiser les données.

Il pourra être envisagé de cataloguer l'enquête par questionnaire. Pour ce faire, et afin d'anonymiser les données, il sera probablement nécessaire d'agréger certaines variables de géolocalisation ou de recoder cette même variable géographique (avec un code géographique de référence plus vaste). Le catalogage de l'enquête qualitative nécessite un temps de réflexion puisqu'il sera peut-être plus difficile d'anonymiser les résultats. Toutefois, au vu du sujet, il semble intéressant d'archiver ces données à des fins de recherches scientifiques longitudinales.

L'équipe va travailler avec les services de la TGIR pour stocker ses données. Il semble donc opportun de mobiliser les outils d'indexation des

données proposés par ces derniers, tel que Nakala par exemple.

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

La base de données produite par questionnaire pourra, selon le poids et le nombre de répondants, être ouverte dans un logiciel de traitements statistiques (SAS ? R ?) ou dans un logiciel type tableur (Excel, Office).

L'équipe doit avancer dans sa recherche pour envisager le partage de l'enquête qualitative et son accessibilité.

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

Il semble pertinent d'attribuer un identifiant unique et pérenne au jeu de données produit via l'enquête par questionnaire. Au sein de l'UMR ESO, des ingénieurs connaissent les démarches pour attribuer ce DOI.

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) ?

Anne Cordier, responsable scientifique du programme, est responsable de la gestion des données.

Le questionnaire est géré via Qualtrics à l'Université de Toulouse, par André Tricot et Anne Cordier. Les données seront ensuite exportées et stockées sur la Box Huma-Num.

Tout ce qui relève de l'enquête qualitative sera saisi par les chercheurs du programme et stocké dans la box HUMA-num.

Stéphane Loret (ingénieur d'études à l'UMR ESO), a ouvert un espace partagé dans la BOX HUMA-Num, Anne Cordier en gère les accès.

Laura Pauchard et Anne Cordier ont élaboré ce plan de gestion de données qui sera examiné par Elie Allouche (DNE). Laura Pauchard et Anne Cordier réaliseront les amendements au plan de gestion de données au cours du programme (espace partagé dans DMP OPIDOR).

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) ?

Le temps alloué à ces questions n'a pas été chiffré et l'équipe ne bénéficie, pour l'instant, pas de moyens supplémentaires pour recruter du personnel.

Anne Cordier s'appuie sur des ressources internes à l'UMR pour avancer sur ces questions (une ingénieure d'études en production traitement analyse de données ainsi qu'un ingénieur d'études CNRS). Nous envisageons aussi de nous appuyer sur les ressources du service de la DRI de l'Université de Caen que nous solliciterons très prochainement.

Le coût de stockage est assuré par la MSHB de Nantes, en lien étroit avec la TGIR HUMA-Num.

Des formations de l'équipe actuelle et le recrutement d'un personnel seraient à envisager pour rendre les données FAIR.
