



HAL
open science

Plan de gestion des données de l'EquipEx+ COMMONS (2.0)

Marie Pellen, Nicolas Larrousse, Dominique Roux, Edith Cannet, Arnaud Daret, Heidi Yaya, H el ene Jouguet, Simon Dumas Primbault, Jean-Fran ois Riv iere, Arnaud Gingold

► To cite this version:

Marie Pellen, Nicolas Larrousse, Dominique Roux, Edith Cannet, Arnaud Daret, et al.. Plan de gestion des donn ees de l'EquipEx+ COMMONS (2.0). Aix-Marseille Universit ; EHESS; Universit  de Caen - Basse Normandie; Universit  d'Avignon et des Pays de Vaucluse; Open Edition. 2023, 16 p. hal-04542731

HAL Id: hal-04542731

<https://hal.science/hal-04542731v1>

Submitted on 11 Apr 2024

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 destin e au d p t et   la diffusion de documents scientifiques de niveau recherche, publi s ou non,  manant des  tablissements d'enseignement et de recherche fran ais ou  trangers, des laboratoires publics ou priv s.



Distributed under a Creative Commons Attribution 4.0 International License

DMP du projet "COntortium of Mutualised Means for OpeN data & Services for SSH"

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 "COntortium of Mutualised Means for OpeN data & Services for SSH"
Version	Version intermédiaire
Objet/périmètre du plan	Le plan couvre le périmètre des données créées uniquement dans le cadre du projet d'EquipeX COMMONS par les trois partenaires impliqués. Il est ainsi indépendant des autres DMP pouvant exister par ailleurs.
Domaines de recherche (selon classification de l'OCDE)	Computer and information sciences, History and archaeology, Art (arts, history of arts, performing arts, music), Social and economic geography, Languages and literature, Philosophy, ethics and religion, Psychology, Political science, Economics and business, Sociology
Langue	fra
Date de création	2022-04-22
Date de dernière modification	2023-11-09
Licence	Creative Commons Attribution 4.0 International

Renseignements sur le projet

Titre du projet	CONsortium of Mutualised Means for OpeN data & Services for SSH
Acronyme	COMMONS
Résumé	<p>COMMONS est un équipement qui construit un environnement numérique unifié accessible au travers d'une offre de services liant données et publications. Il s'inscrit pleinement dans les logiques et dans les objectifs de mise en œuvre des principes de la science ouverte dans le domaine des SHS. Il optimise par une meilleure articulation et une plus grande interopérabilité les services conçus dans le cadre de collaborations déjà en oeuvre entre les trois infrastructures Huma-Num, Métopes et OpenEdition et par la création des services nouveaux afin d'en intensifier les usages en simplifiant l'appropriation par l'utilisateur de cet environnement. L'équipement couvre les besoins de l'ensemble du cycle de vie d'un projet de recherche de la collecte de données à la publication scientifique en procurant les différents outils et services numériques permettant la mise en œuvre des principes FAIR. COMMONS est pensé pour transformer en profondeur les pratiques inscrites dans le cycle de la recherche et leur impact sociétal. En s'inscrivant dans un écosystème standardisé, il augmente la visibilité internationale des SHS françaises. Au cours de l'année 2022, le projet s'est concentré sur la modernisation de l'infrastructure logicielle des trois infrastructures (Huma-Num, Métopes et OpenEdition), cette phase passera en production en 2023 et permettra ensuite d'optimiser l'ensemble des services et de déployer les innovations.</p>

Sources de financement

- Agence Nationale de la Recherche : 21-ESRE-0045

Date de début 2021-11-15

Date de fin 2029-06-22

Partenaires

- Open Edition Center (201420921J)
- Humanités Numériques (201320910B)
- Ecole des Hautes Etudes en Sciences Sociales ()
- Université d'Aix-Marseille ()
- Université de Caen Normandie ()
- Université d'Avignon et Pays du Vaucluse ()
- METOPES (201221521V)

Produits de recherche :

1. Modules d'interconnexion des infrastructures et logiciels d'édition (Logiciel)
2. Documentation de suivi et de gestion du projet (Texte)
3. Documentation technique des logiciels (Texte)
4. Observatoire des pratiques (Jeu de données)
5. Matériel et guides pédagogique (Ressource interactive)

Contributeurs

Nom	Affiliation	Rôles
Pellen Marie - 0000-0002-0333- 7923	Open Edition Center - CNRS	<ul style="list-style-type: none">• Personne contact pour les données (Obs des pratiques, Modules et logiciels, Documentation tech, Matériel pédagogique, Documentation projet)• Responsable du plan de gestion de données

DMP du projet "COnsortium of Mutualised Means for Open data & Services for SSH"

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

Modules d'interconnexion des infrastructures et logiciels d'édition

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

Les logiciels sont produits par des développeurs informatiques appartenant aux unités participant au projet ou par des prestataires.

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

Il s'agit de la rédaction de code pour des développements logiciels. Les principaux types de données générés : code, exécutables

Documentation de suivi et de gestion du projet

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

La documentation du projet est rédigée par l'ensemble des participants au projet et est centralisée par la personne chargée de la coordination du projet.

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

Il s'agit des documents permettant d'assurer le suivi et la gestion du projet (plannings, comptes-rendus, présentations, documents de travail etc). Les principaux types de données : documents bureautiques

Documentation technique des logiciels

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

La documentation technique est rédigée par les membres des unités participant au projet.

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

Il s'agit de la rédaction de la documentation technique des différentes briques logicielles développées. Ces documentations permettront la reprise et la poursuite des développements. Les principaux types de données générés : codes, ODD, documents bureautiques, documentation en ligne

Observatoire des pratiques

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

Des données préexistantes, générées par le fonctionnement normal des infrastructures, seront récoltées : logs serveurs.
Des données nouvelles seront produites : questionnaires, sondages, entretiens, notes de terrain.

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

L'observatoire des usages permettra de mettre en œuvre les outils pour comprendre et analyser l'utilisation de l'équipement en vue d'évaluer son usage et l'adapter le cas échéant. La nature des données utilisées va se préciser au cours du développement du projet. Données pressenties :

1. Logs des serveurs des 4 plateformes d'OpenEdition, d'Isidore et de Nakala
Nature : texte (format générique : IP - Timestamp - Requête HTTP)
Format de fichier : .txt

Volumétrie : environ 800Go / an en compressé pour tout OpenEdition ; Moins de 500 Gos pour la volumétrie des logs de ISIDORE et NAKALA

Sensibilité : IPs en clair, en accord avec la politique de confidentialité d'OpenEdition et annoncé sur le site. Les IPs des logs de NAKALA & ISIDORE sont également accessibles et utilisées avec des outils comme Matomo, mais il est prévu d'en développer les usages dans le cadre de l'observatoire.

1. Métadonnées des contenus publiés / exposés sur les 4 plateformes d'OpenEdition, d'Isidore et de Nakala
2. Réponses à des questionnaires à choix multiples
Nature : texte
Format de fichier : .csv
Volumétrie à préciser (quelques milliers d'utilisateurs)
Sensibilité : utilisateurs pseudonymisés grâce à une table de correspondance
3. Enregistrements d'entretiens semi-directifs
Nature : audio (voire vidéo) et transcription textuelle
Formats de fichier : .mp3, .mp4, .txt
Volumétrie à préciser (quelques dizaines d'utilisateurs)
Sensibilité : transcriptions pseudonymisées grâce à une table de correspondance
4. Réponses à des formulaires d'auto-évaluation avant / après une formation
5. Notes d'observation in situ

Matériel et guides pédagogiques

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

Le matériel pédagogique est réalisé par les membres des unités participant au projet et sera intégré à un système de gestion des apprentissages (Moodle)

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

Création et rédaction de supports de formations pédagogiques pour l'utilisation de l'équipement.

Les principaux types de données générés :

- Supports pédagogiques écrits
- Vidéos
- Contenus interactifs diffusables
- Evaluations
- Parcours de formation

2. Documentation et qualité des données

Modules d'interconnexion des infrastructures et logiciels d'édition

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

Le versionning avec la technologie Git permet d'identifier pour chaque ligne de code produite, son historique de version avec identification du ou des contributeurs (nom et adresse e-mail et la date de chaque modification).

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

Le code est documenté en son sein selon les bonnes pratiques en vigueur

Un ODD sera établi pour les schémas XML-TEI.

Relectures croisées du code par d'autres membres du pôle développement

OpenEdition, Huma-Num et Métopes : Analyse statique de code (SonarQube ou équivalent)

OpenEdition, Huma-Num et Métopes : Ecriture de tests unitaires et fonctionnels par les développeurs

OpenEdition, Huma-Num et Métopes : Mise en place d'une système d'intégration continue dans Gitlab permettant l'exécution systématique des tests pour éviter les régressions.

Documentation de suivi et de gestion du projet

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

La documentation du projet est versionnée et stockée dans le dépôt ShareDocs (proposé par Huma-Num) avec des métadonnées permettant la recherche

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

Les documents sont publiés sous la responsabilité scientifique du projet et validés par les différents acteurs.

Documentation technique des logiciels

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

La documentation technique est versionnée avec la technologie Git. Il est donc possible comme pour les logiciels d'identifier pour chaque ligne, son historique de version avec identification du ou des contributeurs et la date de chaque modification.

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

Relecture croisée par plusieurs membres du projet des documentations rédigées

Observatoire des pratiques

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

1. Logs : un fichier readme accompagnera chaque livrée de logs pour en spécifier : les opérations de pré-processing (notamment ezParse), le format des logs (typiquement IP - timestamp - requête HTTP), d'éventuels commentaires (ex. une interruption de service)
2. Métadonnées des contenus publiés / exposés
3. Réponses à des questionnaires à choix multiples : les données de réponse seront accompagnées d'un fichier readme contenant le questionnaire, des dates d'ouverture du questionnaire, des voies de diffusion du questionnaire
4. Enregistrements d'entretiens semi-directifs : les enregistrements audiovisuels et leurs transcriptions seront accompagnés d'un fichier readme contenant la trame d'entretien, le codage utilisé pour analyser les transcription, ainsi qu'un tableau récapitulatif des pseudonymes, de leurs profils de base (âge, profession), de la durée de l'entretien, de la date de l'entretien ; une table de correspondance des pseudonymes avec les identités sera stockée en local
5. Réponses à des formulaires d'autoévaluation : les réponses aux formulaires seront accompagnées d'une fichier readme contenant le formulaire, la date, le lieu et le contenu de la formation, un tableau récapitulatif des pseudonymes, de leurs profils de base (âge, profession), de la durée de l'entretien, de la date de l'entretien ; une table de correspondance des pseudonymes avec les identités sera stockée en local
6. Notes d'observation in situ

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

Vérification croisée par le PI et les collaborateurs de l'observatoire.

Matériel et guides pédagogique

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

- Documentation technopédagogique pour le formateur intégrant un volet Données sur les systèmes et outils utilisés (Moodle / H5P)
- Métadonnées de description du matériel pédagogique conçus à partir des outils de création et du système des gestion des apprentissage utilisés.
- Données descriptives des évaluations proposées aux stagiaires via les outils de diffusion des contenus pédagogiques

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

- Collaboration ingénieur pédagogique / Expert-e-s pour la validité de la conception pédagogique
- Relecture croisée par plusieurs membres du projet des livrables
- Tests des livrables mis à en oeuvre sur Moodle

3. Stockage et sauvegarde pendant le processus de recherche

Modules d'interconnexion des infrastructures et logiciels d'édition

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

Le code développé est versionné sur une forge (au moins un serveur Gitlab d'une des trois infrastructures).

Les dépôts Gitlab d'OpenEdition et d'Huma Num sont sauvegardés selon les modalités de sauvegarde proposées par l'hébergeur (Centre de calcul de l'IN2P3) : IBM Spectrum Protect. <https://doc.cc.in2p3.fr/fr/Data-storage/mass-storage/backup.html>

Pour Métopes, le dépôt Gitlab de l'université de Caen Normandie est sauvegardé via un CEPH mutualisé fourni par Normandie Université (Université de Rouen Normandie et Université de Caen Normandie)

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

- La sauvegarde des données est décrite au point 3.a.
- Pas de données sensibles.

Documentation de suivi et de gestion du projet

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

Les documents de suivi et de gestion du projet sont stockés sur un service sécurisé de l'ESR permettant le travail collaboratif ShareDocs (Huma-Num).

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

recherche ?

- La sauvegarde des données est décrite au point 3.a.
- Pas de données sensibles.

Documentation technique des logiciels

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

Le documentation technique est versionnée avec la technologie Git et sauvegardée selon les mêmes modalités que les logiciels.

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

- La sauvegarde des données est décrite au point 3.a.
- Pas de données sensibles.

Observatoire des pratiques

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

Disque local à OpenEdition sauvegardé.

Exposition des transcriptions d'entretiens / réponse aux questionnaires anonymisés sur Nakala.

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

Les données ne seront accessibles qu'aux collaborateurs de l'observatoire des usages (OpenEdition Lab et deux partenaires universitaires) par connexion locale au réseau d'OpenEdition.

Les tables de correspondance des identités ne seront accessibles qu'au PI et ses collaborateurs.

Les enquêtés soumis à des questionnaires ou entretiens recevront une fiche d'information et retourneront un formulaire de consentement signé.

Matériel et guides pédagogique

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

Les documents de suivi et de gestion du projet sont stockés sur un service sécurisé de l'ESR permettant le travail collaboratif ShareDocs (Huma-Num).

Les données sont stockés sur un système de gestion des apprentissages Moodle dont l'hébergement est confié à un prestataire (E-Learning Touch)

Un système de sauvegarde des données est mis en place par le prestataire.

Ces données sont accessibles sous conditions d'accès aux parcours de formation (compte utilisateur, inscription à la formation).

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

- La sauvegarde des données est décrite au point 3.a.
- Pas de données sensibles.

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 ?

Conformément à l'article 37 du RGPD, une Déléguée à la Protection des Données (DPD) a été désignée par Huma-Num et OpenEdition. Le service de protection des données personnelles du CNRS et notamment sa DPD travaille en collaboration avec les différentes structures et en soutien de ces dernières sur les questions liées aux RGPD afin d'assurer la conformité de ces établissements et de certains de ses utilisateurs (chercheurs) dans ce domaine.

Pour OpenEdition :

- La Politique de confidentialité traitant de la protection des données à caractère personnel des internautes/utilisateurs sur le portail d'OpenEdition a été mise à jour.
- Une annexe "Données Personnelles" destinées aux Editeurs pour les plateformes d'OpenEdition Journals, OpenEdition Books et d'Hypothèses sera visible à la suite des nouvelles CGU destinées aux Editeurs.
- Une annexe "Données Personnelles" destinées aux Annonceurs de Calenda est déjà visible à la suite des nouvelles CGU destinés aux Annonceurs : <https://calenda.org/1044726>.

La chargée des affaires juridiques coordonne la rédaction et la mise à jour de ces documents juridiques produits. Elle est l'intermédiaire entre les différents interlocuteurs notamment entre :

- les équipes opérationnels internes et la DAJ du CNRS
- les équipes opérationnelles internes (et potentiel partenaire externe) et ProtisValor Méditerranée, filiale de valorisation de la recherche d'Aix-Marseille Université.

Pour Huma-Num

- Les modalités de gestion des données personnelles pour Huma-num sont disponibles sur la page : <https://humanid.huma-num.fr/>

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 ?

Législations françaises applicables :

- les dispositions du code de propriété intellectuelle relative à la propriété littéraire et artistique
- la loi n° 2004-575 du 21 juin 2004 pour la confiance dans l'économie numérique (LCEN)
- la loi du 29 juillet 1881 modifiée sur la liberté de la presse,
- la loi n°78-17 du 6 janvier 1978 modifiée relative à l'informatique, aux fichiers et aux libertés
- la législation relative aux atteintes aux systèmes de traitement automatisé de données (art. L323-1 et suivants du code pénal)
- la loi n° 94-665 du 4 août 1994 modifiée relative à l'emploi de la langue française
- la loi n° 2016-1321 du 7 octobre 2016 pour une République numérique

I. Concernant les droits de propriété intellectuelle,

Le choix d'une politique de licences Creative Commons a été convenu sur les plateformes OE. Le déploiement de cette politique est prévu selon

l'ordre de priorité suivant :

- 1) OpenEdition Journals tout d'abord ;
- 2) Un travail similaire sera effectué auprès des éditeurs, une fois que le déploiement de la politique de licence CC sera effectué sur OpenEdition Journals.
- 3) Hypothèses

Les métadonnées de l'ensemble des plateformes d'OpenEdition sont sous mention CC0. [L'information](#) a été transmise aux utilisateurs.

Pour OpenEdition Journals :

Des lettres de confirmation du choix de licence CC pour chacune des revues actives disponibles sur la plateforme OEJ ont été envoyées. En cas d'interrogation des membres des équipes des revues sur les licences CC, des réponses et des informations juridiques leur sont apportées.

De plus, la transmission des bonnes pratiques et des obligations légales est également assurée notamment en leur rappelant : l'obligation d'avoir l'accord préalable des auteurs pour toute mise en place d'une licence CC et l'importance d'avoir des contrats avec les auteurs : renvoi par exemple aux contrats types de Science Ouverte faisant référence aux licences CC.

Résultats : A ce jour, sur 617 revues présentes sur OpenEdition Journals :

- Environ 470 revues ont reçu une lettre (soit pour leur rappeler la licence déjà mise en place soit pour leur confirmer leur choix et/ou leur demander des détails d'application sur la licence CC choisie)
- A ce jour, environ 431 revues ont une licence CC appliquée sur leur site
- Une soixantaine de revues sont inactives ou éteintes
- 97 revues n'ayant pas déclaré de licence cc : une relance est en cours pour ces revues restantes.

Documentation/ressources :

- Tableau général de suivi avec le choix des licences CC par revue présentes sur OEJ
- Page dédiée à la politique de licences CC sur OpenEdition : <https://www.openedition.org/40528>

Les droits de propriété intellectuelle sont mentionnés de façon plus détaillée dans les CGU des plateformes/services impliqués dans le projet.

II. Concernant la conformité à la LCEN

Objectif : La création, la mise à jour et la mise en ligne des mentions légales sont destinées à mettre en conformité la structure en prévoyant les éléments d'identification obligatoire imposés par l'article 6 III de la loi du 21 juin 2004 pour la confiance dans l'économie numérique (LCEN).

Résultats ou Réalisation à venir :

Pour OpenEdition

- La rédaction des mentions légales a été faite et validées par la Direction des affaires juridiques du CNRS pour la version française. S'agissant de la version anglaise, celle-ci a été soumise à une juriste assermentée dans la traduction juridique.
- Réalisés : Les pages des Mentions légales sur les portails OEJ, OEB, CALENDAL ont été créées par le service Données.
- A venir : La création de la page des Mentions légales pour Hypothèses est mise en attente de la mise en place de LODEL2.

Pour Huma-Num

- Mise à jour des mentions légales de Nakala – en cours
- Mise à jour des mentions légales d'Isidore – en cours
- Si nécessaire, mise à jour des CGU de la « grille de services » d'Huma-Num disponibles : <https://documentation.huma-num.fr/cgu/>
 - Dans les CGU : mentionner le rôle d'hébergeur au sens de la LCEN

Pour METOPES

- Mise à jour des mentions légales en cours

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

Concernant la dimension éthique et l'intégrité scientifique, l'Inist-CNRS rappelle qu'« Etre informé de l'évolution des règles juridiques est aussi très important pour un partage responsable des données utilisées. ».[1]

De plus, le [code de conduite européen pour l'intégrité en recherche](#) identifie quatre valeurs fondamentales : responsabilité, respect, honnêteté et fiabilité.

Les CGU contribuent à la régularité du fonctionnement des instances statutaires des établissements. De plus, en encadrant les relations entre les plateformes (et les structures qui les gèrent) et leurs usagers/utilisateurs, elles garantissent une transparence et une information loyale et fiable sur les services proposés.

Pour les plateformes d'OE, une refonte des CGU est en cours. L'ordre de priorité est le suivant :

1/ OpenEdition Journals et OpenEdition Books; 2/ Hypothèses.

Les propositions des nouvelles versions des CGU pour OEJ et OEB sont transmises à la DAJ du CNRS et soumises à validation d'un cabinet d'avocat (réunion prévue le 28/09/2023).

La refonte des CGU de CALENDAL est faite depuis le 8 mars 2023. Les CGU sont en ligne :

- o CGU destinées aux internautes : <https://calenda.org/1044719>

o CGU destinées aux annonceurs : <https://calenda.org/1044726> .

Pour Huma-Num :

- Une première version des CGU de Nakala est en cours de rédaction.
- Une version des CGU d'Isidore est en cours de révision.

Dans le cadre de la politique de licences CC sur les plateformes OE :

- L'information sur les bonnes pratiques et des obligations légales est également assurée notamment en leur rappelant : l'obligation d'avoir l'accord préalable des auteurs pour toute mise en place d'une licence CC et l'importance d'avoir des contrats avec les auteurs : renvoi par exemple aux [modèles de contrats de Science Ouverte](#) faisant référence aux licences CC.

- Pages d'informations sur les licences CC : <https://www.openedition.org/40395> recensant notamment des sources d'informations complémentaires en bas de page à ce sujet ;

- [FAQ Contrats et responsabilité légale](#) sur les contenus publiés sur OpenEdition Books et OpenEdition Journals notamment portant sur les droits de diffusion / contrat ;

- En cours : tableau synthétique des droits et obligations des utilisateurs en fonction des licences CC afin de traduire de manière claire et synthétique les usages possibles.

III. Prospectif

L'idéal serait de faire en état des besoins juridiques en fonction de chaque produit de la recherche créé. Voici un début :

- o Modules logiciels : choix d'une licence libre par exemple pour Lodel 2
- o Documentation projet : pas vocation à être diffusée au public (seulement aux instances et interne aux trois structures du projet)
- o Documentation technique : a priori ni PI ni données personnelles
- o Observatoire des usages : attention toute particulière aux données personnelles des usagers
- o Matériel pédagogique : il est prévu d'opter pour une licence CC pour les formations mises à disposition dans Moodle + documentation tuto formation

Veille juridique :

Le [règlement](#) (UE) 2022/2065 du Parlement européen et du Conseil du 19 octobre 2022 relatif à un marché unique des services numériques et modifiant la directive 2000/31/CE (règlement sur les services numériques) entre en vigueur le 17 février 2024.

Les services « intermédiaires » à des destinataires se situant au sein de l'Union européenne (notamment les plateformes et les moteurs de recherche) devront se conformer aux nouvelles obligations issues de ce règlement "Digital Services Act".

Pour METOPES :

CGU en cours de rédaction

5. Partage des données et conservation à long terme

Modules d'interconnexion des infrastructures et logiciels d'édition

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 ?

Autant que possible, les éléments logiciels sont publiés sous licence libre (de type GPL, CECIL ou BSD qui seront à préciser selon le type de développement). Certains éléments logiciels intégrant un logiciel existant propriétaire ou spécifique à l'environnement d'une plateforme ne pourront être publiés sous licence libre.

Le code publié sous licence libre sera accessible sur le dépôt Gitlab d'au moins une des trois infrastructures et pour certains logiciels sur Github

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) ?

Non défini pour le moment.

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

Le code publié sera accessible via Git (Gitlab ou Github) et le cas échéant dans des dépôts de paquets Python (<https://pypi.org>) ou PHP (<https://packagist.org>)

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

Non défini pour le moment.

Documentation de suivi et de gestion du projet

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 ?

Sans objet, ces documents ne sont pas destinés à être partagés.

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) ?

Sans objet.

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

Aucun outils ni méthodes particuliers

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

Sans objet sauf si l'on décidait de publier la documentation

Documentation technique des logiciels

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 ?

La documentation des logiciels sous licence libre sera diffusée sous licence Creative Commons (CC By-NC-SA <https://creativecommons.org/licenses/by-nc-sa/4.0/> ou autre licence Creative Commons) .

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) ?

Non défini pour le moment.

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

Le code publié sera accessible via les technologies Git (Gitlab ou Github).

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

Non défini pour le moment.

Observatoire des pratiques

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 ?

Avec l'accord des enquêtés, les transcriptions anonymisées des entretiens et les réponses anonymisées aux questionnaires ou formulaires seront mises en ligne avec le protocole d'enquête sur Nakala.

Les logs d'OpenEdition seront disponibles en accès restreint avec les IPs hachées sur demande à l'équipe d'OpenEdition.

Le protocole d'enquête final sera publié sur Zenodo.

Des publications de recherche seront publiées dans des revues nationales et internationales à comité de lecture en informatique, science des données, sciences de l'information et de la communication, sociologie.

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) ?

Avec l'accord des enquêtés, les transcriptions anonymisées des entretiens seront mises en ligne avec le protocole d'enquête sur Nakala.

Les logs d'OpenEdition seront disponibles en accès restreint avec les IPs hachées sur demande à l'équipe d'OpenEdition.

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

Le code publié sera accessible via les technologies Git.

Les résultats de l'observatoire seront accessibles via une interface web classique.

Lecture des transcriptions des entretiens : lecteur .txt.

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

DOI pour les données exposées et déposées dans NAKALA

Identifiant de Software Heritage pour le code publié sur git.

Pas d'identifiant pour les logs disponibles sur demande.

Matériel et guides pédagogique

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 ?

Les conditions particulières de partage des données du matériel pédagogique dépendent du lieu de diffusion, du type de matériel, des conditions d'utilisation fixées qui peuvent varier.

Matériel pédagogique hébergé sur le système de gestion des apprentissages Moodle d'OpenEdition

Types de matériel : vidéos, objets d'apprentissage interactifs, fichiers d'exercice, glossaires, évaluations, scénarios pédagogiques

Types d'accès : l'utilisateur dispose d'un compte sur Moodle et son inscription à une formation est validée par OpenEdition ou un partenaire

Sauf mention contraire au niveau de chaque unité de matériel pédagogique, une licence Creative Commons CC-BY est appliquée par défaut à l'ensemble des contenus d'une formation hébergée sur Moodle, avec la possibilité d'appliquer des exceptions de licences à chaque unité de matériel pédagogique (ex : un fichier, une vidéo). Cette mention est affichée en introduction des formations. Quand cela est possible, la licence est embarquée en tant que métadonnée dans les différents fichiers qui constituent le matériel pédagogique. Le cas d'usage le plus courant est celui des objets d'apprentissages créés à partir de l'outil-auteur H5P qui permet de leur affecter une licence et de l'afficher, de renseigner des métadonnées, d'ajouter des boutons de téléchargement pour faciliter leur réutilisation dans des environnements de diffusion compatibles (principalement LMS).

Les exceptions de licence prévues dépendent essentiellement de la nature du contenu et sont fixées par les producteurs de contenus en accord avec la direction. Une licence plus fermée peut être préférée afin de préserver l'intégrité d'un contenu pédagogique et du message qu'il contient.

Il est prévu que la totalité d'un cours puisse être téléchargeable sous forme d'archive pour permettre à un utilisateur (ex : formateur, ingénieur pédagogique) de l'importer dans un environnement d'apprentissages capable de traiter ce type d'archive.

Il est possible qu'une formation (ex : parcours d'auto-formation en ligne) et les contenus qu'elle contient soit ouverte sans mécanisme de validation de demande d'inscription par OpenEdition.

Matériel pédagogique hébergé en dehors du système de gestion des apprentissages Moodle d'OpenEdition

Types de contenus : vidéos, objets d'apprentissage interactifs, fichiers d'exercice diffusés à l'extérieur du système de gestion des apprentissages

Types d'accès : page web, pas de condition d'accès

Certains peuvent être proposés à la consultation sans contrainte d'accès pour faciliter leur découverte et leur réutilisation : par exemple le cas des contenus hébergés sur Nakala ([exemple 1](#), [exemple 2](#)).

Pour des raisons de cohérence, ces contenus partagent les mêmes conditions d'utilisation que ceux hébergés sur Moodle.

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 qui constituent le matériel pédagogique ne sont pas toutes conservées de manière pérenne.

Plusieurs raisons peuvent expliquer cela : complexité de mise en œuvre (ex : hébergement pérenne de la plateforme Moodle), absence de services adaptés au type de contenu pédagogique à héberger.

Certains contenus du matériel pédagogique sont toutefois compatibles avec un dépôt en archive ouverte. Ce mode d'hébergement est privilégié pour les contenus qui répondent à ce critère, en utilisant par exemple Nakala.

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

Pour accéder aux ressources et les télécharger :

- navigateur web pour les ressources publiques,
- inscription à un parcours de formation sur Moodle pour les ressources intégrées à ces parcours

Pour éditer les ressources :

- Logiciel de montage et d'édition vidéo Camtasia
- H5P ou H5P.com pour le matériel pédagogique interactif

Pour diffuser les ressources :

- iFrame pour les ressources embarquables accessibles sans restriction
- LMS avec des capacités d'import d'archives de exportés depuis Moodle

- Environnement de diffusion interopérable avec H5P

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

Tout contenu intégré à Nakala disposera d'un identifiant de type DOI publié.

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) ?

Marie Pellen
Responsable du projet
OpenEdition CNRS / AMU / EHESS / UA

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 projet est doté d'un référent FAIR dont la mission est transversale aux différents éléments du projet et a pour but de définir un processus de FAIRisation des données produites par le projet. Le processus prendra la forme d'une évaluation continue placée sous la responsabilité scientifique du projet.

Un comité FAIR du projet COMMONS sera constitué pour formuler des propositions et s'assurer de l'optimisation de l'intégration des principes FAIR tout au long de la mise en place de l'équipement, tant pour les données produites que pour celles qui seront hébergées. Le comité examinera des implémentations des principes FAIR précédemment réalisées au sein des infrastructures et communiquera des principes méthodologiques de FAIRisation partagés par la communauté scientifique internationale (p. ex. : grille d'analyse FIP (FAIR Implementation Profile - <https://www.go-fair.org/how-to-go-fair/fair-implementation-profile/>)). Le partage de ces méthodes d'auto-évaluation de la conformité aux principes FAIR permettra de construire la stratégie de FAIRisation du projet.