

Séminaire science ouverte et données – HERMES Spectacle vivant

22 janvier 2026 – Jeanne Fras, data ingénieure et Clarisse Bardiot, PU Études
théâtrales en charge de l'ERC STAGE

RAPPELS

- séminaire en présentiel sur le livre blanc **lundi 27 avril à Paris**
- demande de Sorbonne Nouvelle d'entrer dans un Plan de Gestion de Données (PGD)
- Pôle ARDOISE Rennes 2 : accompagnement en gestion de données

SCIENCE OUVERTE ET ETHIQUE DE LA RECHERCHE POUR LES ARTS DU

SPECTACLE : RETOUR D'EXPERIENCE DU PROJET STAGE

ERC STAGE 2024-2028 : recherche sur l'écriture de l'histoire du théâtre dans le contexte d'archives qui sont des traces numériques :

- méthodologie : comment mobiliser ces archives ?
- épistémologie : quel type d'histoire peut-on écrire ?

⇒ terrain de recherche : le Festival d'Avignon, pour en écrire une histoire du point de vue des équipes artistiques et techniques (et non directions/ publics/ œuvres comme habituellement)

⇒ triple corpus :

1. programmes généraux et programmes de salles —> extraction automatisée des informations pour comprendre l'évolution des carrières, des métiers techniques, etc.
2. photographies (notamment de Raynaud de Lage) et captations vidéo —> analyse via algorithme de vision par ordinateur, sur des questions esthétiques de réminiscence entre les œuvres
3. processus de création —> au-delà de l'observation des répétitions, prise en compte de l'ensemble des traces numériques du processus dans l'ensemble du projet et de la compagnie (donc récupération des espaces de stockage numériques pour collecte des traces)

⇒ points communs avec HERMES : collecte et stockage de données ; nécessité de les rendre disponibles au-delà du projet et de travailler dessus à plusieurs

⇒ décision de recruter un ingénieur spécialisé dans la gestion des données pour l'ensemble du projet ERC

GESTION DES DONNEES

DEFINITIONS

- données de la recherche : fait ou chiffre collecté pour être analysé et traité pour alimenter réflexions, discussions ou calculs —> tout ce qui est collecté dans le cadre de la recherche
- partage et réutilisation *possibles dans le cadre de la science ouverte* : données brutes, données traitées, données analysées, corpus, jeux de données, publications
- Plan de Gestion de Données : document explicitant le cadrage du projet et la collecte et stockage des données dans ce cadre et au-delà du projet :
 - cadre
 - collecte et réutilisation
 - documentation des données
 - stockage
 - accès et partage
 - publications
 - archivage et préservation à long terme (pérennité des données)
 - coûts
 - auto-évaluation éthique et juridique

BIEN GERER SES DONNEES PENDANT LA RECHERCHE

- *comment collecte-t-on les données ? avec quels moyens ? dans quel cadre ?*
- *quelles données sont collectées ?*
- *comment les stocke-t-on et les organise-t-on*
- *comment les archive-t-on ?*

⇒ prévoir des endroits de sauvegarde (3 différents) tels que :

- disque dur
- cloud privé : OVH, etc.
- outil institutionnel : Koumoul, Huma-Num, ShareDocs

⇒ organiser et normaliser les données et le corpus général dans une arborescence uniforme facile à explorer et partager :

- employer une charte de nommage commune

PARTAGER SELON LES PRINCIPES FAIR

F = Facilement Découvrable

A = Accessible
I = Interopérable
R = Réutilisable

⇒ ces principes permettent un accès et une préservation à long terme, notamment grâce à un DOI (adresse pérenne) et à un format ouvert ⇒ il faut anticiper les conditions de réutilisation et les droits associés

L'interopérabilité nécessite l'usage de formats ouverts et non propriétaires : l'outil Facile du CINES permet de vérifier la compatibilité des formats envisagés.

STOCKER ET PUBLIER LES DONNEES

⇒ privilégier des entrepôts ouverts selon la thématique de recherche : Data.gouv, Progedo, Nakala ⇒ privilégier la publication sur HAL
⇒ ces plateformes offrent un numéro de DOI nécessaire à la pérennité
⇒ DMP Opidor permet une rédaction et publication facilitée du PGD

AUTO-EVALUATION ETHIQUE ET JURIDIQUE

Comment respecter et protéger le droit individuel dans le cadre de l'ouverture des données ?

Éléments à prendre en compte

⇒ identifier les données personnelles : elles permettent une identification directe par une seule donnée (ou indirecte, en croisant plusieurs données) de la personne concernée —> nom, voix, etc.
⇒ identifier les données sensibles : opinions, sujets de santé, etc. ⇒ identifier ce qui relève du droit d'auteur

Les espaces de stockage choisis doivent être en capacité de protéger les données personnelles et sensibles qui y sont stockées.

Situations à risque

⇒ publication des données
⇒ fuite de données

Interlocuteurs

⇒ CNIL : anonymisation, pseudonymisation et application du droit de retrait des données
⇒ RGPD : protection des données personnelles ; fournit des formulaires de consentement, des conventions et des documents d'information
⇒ AI Act : encadre l'usage des données par IA au niveau européen

Solutions possibles

⇒ anticiper les risques et les impacts
⇒ anonymisation ou pseudonymisation des données
⇒ recueil du consentement au traitement des données pour la recherche et la publication

- ⇒ recours à une consultation externe spécialisée : juriste, comité d'éthique, etc.
- ⇒ limitation des publications ou publication sous condition

QUESTIONS

Marion précise que nos travaux entrent dans le PGD global d'HERMES mis en place par l'Université Sorbonne Nouvelle ⇒ allègement de la charge sur ces questions mais nécessité de réfléchir à l'harmonisation de nos pratiques.

QUESTIONS SUR LES FORMATS

Marion : la suite Office fait-elle partie des formats ouverts ?

Jeanne ⇒ il semblerait, les formats .xls, .ppt et .doc sont disponibles

Clarisse ⇒ mais cela reste des formats propriétaires, le vrai format ouvert est la suite LibreOffice

Marion : les documents produits ont-ils intérêt à être stockés en format LibreOffice ?

Jeanne ⇒ oui, c'est l'idéal et ce sera plus robuste

Marion : un document saisi par HAL est visible par tout le monde

Jeanne ⇒ on peut simplement signaler son existence, il existe plusieurs formats de saisie

Clarisse ⇒ il existe différents niveaux de visibilité, par exemple seulement le résumé mais la bonne pratique reste de mettre l'article complet en pré-publication ; certains éditeurs utilisent des formats ouverts tels qu'OpenEdition donc on peut simplement faire le lien sur notre CV HAL

RECUEILLIR LES DONNEES EN ENTRETEN

Clélia : concernant les entretiens, doit-on réfléchir à un formulaire commun en amont des collectes ?

Clarisse ⇒ beaucoup d'entretiens pour les directions techniques, donc ce serait utile et il existe déjà un template lié à STAGE ; importance de l'information des personnes interrogées sur ce en quoi consiste le projet, leurs droits dans ce cadre (retirer la publication/ l'archivage de l'entretien), et ce que l'on fera avec leurs données. Difficile de savoir exactement ce que l'on souhaite faire, on peut faire des entretiens test sans savoir exactement ce que sera la publication issue de ce travail et essayer de prévoir le maximum de possibilités dans les formulaires de consentement.

Clélia : dans le cas de présentation spontanée auprès des personnes interrogées (spectateurices à l'issue d'une représentation), quelle procédure ?

Clarisse ⇒ présenter le formulaire écrit et la notice d'information

Jeanne ⇒ le consentement doit être éclairé, ce que permet cette notice

Marion : ces traces écrites permettent de se protéger en cas de difficultés par la suite

Jeanne ⇒ cela permet normalement qu'il n'y ait pas de changement d'avis

Sophie : que fait-on des archives de ces formulaires pour en assurer la veille et la pérennité ? où et comment les stocker ? Marion : peut-on les stocker également sur Huma-Num ?

Clarisse ⇒ on peut

Jeanne ⇒ on peut scinder en deux espaces de stockage, le projet sur un serveur et la partie administrative sur un autre

Marion : pour l'avenir, le consentement ne couvrira pas la réutilisation des données dans un autre cadre, l'archivage des formulaires ne peut pas permettre cela ?

Jeanne ⇒ cela peut faire partie des conditions (science ouverte) pour le réemploi des données ; la donnée est d'ordre public à partir du moment où le jeu de données est accessible

Clarisse ⇒ cela nous arrive de réutiliser des entretiens faits par d'autres et disponibles dans les archives

Florence : sur la question du TEJ, il est envisagé une sorte de micro-trottoir pour recueillir des paroles d'enfants ; on compte demander l'autorisation des parents, comment faire ?

Jeanne ⇒ il faut le consentement éclairé des parents/ responsables légaux mais aussi de l'enfant ; se questionner sur la nécessité d'avoir l'identité de l'enfant —> pas de donnée sensible si l'enfant n'est pas identifié par son nom/ prénom

Clarisse ⇒ il faut un formulaire un peu spécifique, différent du formulaire général pour les adultes ; dans le cas des mineurs, on doit garantir l'anonymisation

Florence : on peut laisser l'âge, mais changer le prénom ?

Jeanne ⇒ on peut utiliser des services qui permettent d'anonymiser les données, il faut voir sur le site de la CNIL

Clarisse souligne le fait que Google Drive, bien qu'efficace pour le travail synchrone, permet l'entraînement de l'IA Gemini de Google : **ne pas l'utiliser pour des données personnelles et sensibles.**

HARMONISATION DES PRATIQUES

Marion : nos ressources/ documents communs à produire assez rapidement sont la charte de nommage, un espace Huma-Num (dont en se posant la question de l'arborescence, par type de document ou d'activité), les formulaires de consentement aux entretiens, un pas-à-pas sur la gestion et le stockage des données. Est-ce qu'une arborescence Huma-Num/ ShareDocs serait suffisante pour nous ? Il nous faut une vigilance particulière car le projet est une ANR

Clarisse ⇒ on utilise pour STAGE l'outil d'IA Whisper pour la transcription des entretiens, cela doit être documenté dans le PGD et précisé dans le formulaire de consentement ; l'intérêt du PGD est de pouvoir être mis à jour pour inclure de nouvelles méthodes

Marion : pour l'arborescence Huma-Num, envisage-t-on plutôt une arborescence par projets/ activités avec les personnes associées pour gérer le dossier de leur projet ou un autre type d'arborescence ? chaque activité a peut-être des besoins différents ? Il faudrait rediscuter cela lors du séminaire

Clarisse ⇒ sur des gros projets, des dossiers par workpackage/ par équipe pour que chacune classe ses documents selon l'arborescence qui lui convient ; il faut se mettre d'accord et tenir le choix dans la durée

Marion : ouverture de dossiers par type d'activité et chacun·e organise dans son dossier-activité sa propre arborescence en fonction des besoins ; certaines activités sont similaires, comme les conférences performées, que l'on retrouve aussi pour le théâtre en amateur : dans ce cas, on pourrait avoir une approche par type d'équipe

Clarisse ⇒ cela pourrait être une mission du post-doc de vérifier le rangement au sein de l'arborescence, de garantir le respect de la charte de nommage ; on peut déjà créer tout ce qui est nécessaire et cela sera repris à son arrivée

A FAIRE

- charte de nommage des documents (*préparée par Marion*)
- pour les personnes concernées, se mettre d'accord sur un formulaire de consentement commun pour les entretiens ⇒ **à présenter et discuter lors du séminaire du 27 avril**
- présentation globale du projet et du PGD
- ouverture d'un espace de stockage Sharedocs sur Huma Num (*gestion provisoire par Marion*)