

# Compte rendu collaboratif de la journée d'échange inter-consortiums d'Huma-Num sur « Les archives numériques de la recherche »

*1er Février 2016, 10h00-17h00 - 84 rue de Grenelle, 75007 Paris*

## **Objectifs de la journée**

A l'issue d'une réunion préparatoire qui a donné lieu des échanges riches et prometteurs, les participants ont pensé à organiser cette journée en 3 temps : les présentations des consortiums / mieux nous connaître ; échanger sur des questions transversales ; voir comment nous pouvons prolonger cette journée à travers des actions communes, des réseaux ou des groupes de travail.

Objectif général : avoir des **échanges communs**, mener des **réflexions transversales**.

## **Présentation de chaque consortium**

- Son organisation, ses objectifs, les données qu'il traite, les livrables.
- Difficultés rencontrées, solutions apportées ou recherchées.
- Effets du consortium sur la communauté scientifique et la formation à la recherche

8 consortiums sont représentés et successivement présentés :

- Archives des mondes contemporains – ARcMC (Serge Wolikow)

Il est constitué d'historiens sociologues, ethnologues, anthropologues. La dimension historique est très importante (19e-20e siècle).

Le consortium associe la réflexion sur les archives **de** la recherche et **pour** la recherche. Il réfléchit sur la chaîne de production numérique, jusqu'à l'indexation des documents. Un des objectifs du consortium est de soutenir une démarche collective et collaborative pour répondre aux problèmes qui découlent de la nature des archives de la période contemporaine.

La mobilisation de métiers et de compétences très différents (archivistes, documentalistes, chercheurs, spécialistes de l'image etc.) est une dimension importante des projets du consortium. Ces personnes sont souvent localisées dans des espaces (bibliothèques, musées, universités) très divers qui fonctionnent de manière séparée. C'est à la fois une richesse et une difficulté.

Actuellement, le consortium associe 14 partenaires de natures différentes : 4 MSH, 6 équipes ou laboratoires de recherche et 4 institutions patrimoniales. L'INA est également un partenaire essentiel, même s'il n'est pas encore membre du consortium.

Les partenaires ne sont pas seulement académiques mais également issus de l'univers de la documentation ou des archives. La question de l'intégration de l'INA se pose sur différents projets intégrés et développés au sein du consortium : archives audiovisuelles, archives de la seconde guerre mondiale, des mouvements sociaux, du monde rural...

Le consortium est organisé en pôles thématiques :

- Archives de la recherche et des chercheurs
- Circulations, migrations
- Monde de l'entreprise et travail industriel
- Mondes ruraux
- Mouvements sociaux et organisations ouvrières
- Pensée critique.

Depuis deux ans organisé un séminaire sur les archives audiovisuelles qui rassemble plusieurs plateformes MSH et d'autres consortiums en vue de la mise au point d'une méthodologie et de livrables concernant les **bonnes pratiques** en matière d'indexation, de description, de conversions de fichiers, etc...

La saisie de métadonnées adaptées, en accord avec la nature des documents numérisés et publiés est une question qui préoccupe le consortium. Indexe-t-on de la même manière une photo, un tract, de la correspondance, etc. Le développement d'une numérisation de haute qualité même sur des documents historiques anciens permet de développer une méthodologie de fouille de données (par exemple sur 5000 brochures ).

Une partie des instruments de recherche est en ligne et consultable sur le Portail Archives Numériques et DONnées de la Recherche (Pandor), hébergé par la MSH Dijon <https://pandor.u-bourgogne.fr/cdc/html>.

La mise en ligne d'archives de chercheurs, individuelles ou collectives, n'a pas encore été organisée, même si c'est une question qui est à l'ordre du jour.

La diversité sinon l'hétérogénéité des corpus, induit une réflexion sur leur présentation en réseau thématique permettant des usages différenciés : celui des chercheurs avancés, des étudiants, du grand public ...

Le consortium a également un Carnet sur Hypothèses, un compte Twitter et un site en cours de réalisation qui donnera l'accès à toutes les réalisations. <http://arcmc.hypotheses.org/>.

- Archipolis (Bernard Denni)

Consortium avec une dimension plus « sciences sociales », au périmètre « moins important », il s'intéresse aux enquêtes qualitatives réalisées par des chercheurs en sciences sociales du politique (enquêtes, discours, entretiens semi-directifs ou non, etc.).

Les éléments collectés et le processus même de collecte des éléments posent divers problèmes :

- Propriété "intellectuelle" : ces recherches sont très souvent collectives ; difficultés au niveau de la propriété effective.
- Confidentialité : les données sont sensibles et ont donc besoin d'anonymisation.

- Le format des données : la masse de documents avec les données préparatoires mais aussi tout ce qui suit la recherche, les analyses...

De plus, le constat (2012) sur les intérêts d'une réutilisation des archives d'enquêtes qualitatives a amené le consortium à réfléchir aux moyens de réutiliser les données : de très nombreuses enquêtes de qualité existent dont beaucoup sont sous-utilisées. Les données sont très fragiles et risquent de disparaître avec le chercheur qui les a collectées ; constat d'une sous-exploitation des corpus, peu de moyens pour leur préservation.

Le consortium a développé des outils et des procédures pour gérer les archives en étant conforme aux exigences méthodologiques de réutilisation et en accord avec les normes et standards.

Quand les outils seront au point, il travaillera à leur diffusion au sein des laboratoires : chercheurs ET personnels.

Un tandem est né :

- ARCHIPOLIS : 9 laboratoires avec le CDSP (Sciences Po) porteur +
- BeQuali (Base de données d'Enquêtes Qualitatives) : qui a pour but la diffusion des matériaux d'enquêtes qualitatives. La banque d'enquêtes fait partie de l'Equipex DIME-SHS.

L'association de ces deux pôles permet d'échanger des données, de les traiter et de les rendre accessibles (Archipolis recense et traite les métadonnées sur les enquêtes, prépare des notices descriptives et BeQuali les met à disposition).

Le consortium a déjà mis en place un réseau avec des relais locaux, et réalise tout un travail de communication au sein des laboratoires.

Le consortium se pose la question de la pérennisation du dispositif et de la consolidation des liens avec les services de documentation et d'archives des institutions de rattachement des laboratoires. Il s'agit d'un dispositif encore fragile (réseau de laboratoires) et face auquel certains chercheurs restent réticents.

L'un des problèmes soulevé et signalé concerne la relation avec les chercheurs (la réutilisation des enquêtes notamment, la fragilité des structures, etc.).

La relation personnelle du chercheur aux données collectées explique sans doute en partie la difficulté que les chercheurs ont à les mettre à disposition des autres, à les partager avec d'autres scientifiques.

<http://archipolis.hypotheses.org/>

- CAHIER (Thomas Lebarbé)

Acronyme = Corpus d'Auteurs pour les Humanités : Informatisation Edition Recherche.

Consortium dédié aux corpus d'auteurs pour les Humanités indépendamment de toutes les disciplines. Il est né en 2011, pour réfléchir aux moyens de l'informatisation (pérennisation, encodage, etc), de l'édition et de la recherche des données.

L'originalité du consortium est qu'il s'agit d'un consortium de projets et non d'institutions car souvent les projets sont nés en marge des laboratoires de rattachement. Relation innée avec les MSH (dont le but est de favoriser l'interdisciplinarité).

Une communauté créée ex nihilo, qui se co-construit, autour de besoins partagés (outils, normes, littérature numérique...). Un réseau de chercheurs qui se posent des questions similaires. Environ 8 adhésions par an.

Au départ 12 projets, aujourd'hui 42, et demain 80 projets... Comment s'organiser ?

Tous les projets sont représentés au Comité de Pilotage (1 représentant par membre). L'entrée de nouveaux projets est décidée à ce moment-là. Il y a 2 CoPil par an (en juin et décembre). Une dizaine de membres dans le bureau, représentant la diversité géographique des partenaires.

Aller au-delà de la simple *XMLisation* des données : la visée du consortium est le partage, la *moissonnabilité*. C'est penser une nouvelle culture du domaine avec le numérique.

Il y a plusieurs groupes de travail au sein du consortium pour répondre collectivement à des problématiques (juridique - problématique des éditeurs, ayants droits -, Web-OAI, crowdsourcing, typologie textuelle, formation et HumaNum, correspondance, chaînes éditoriales, outils et standards - formations -, méthodologie de recherche, etc.).

Le consortium organise aussi des journées de travail et une semaine annuelle de formation (ex : comment j'encode mes textes, les standards...).

Le consortium a une vocation nationale et compte sur un effet "boule de neige" pour se faire (re)connaître dans une communauté qui ne se connaît pas.

Constat de la même difficulté des chercheurs à partager les sources produites.

Difficultés pour planifier des formations, des actions, etc. Difficultés aussi pour faire reconnaître le modèle de fonctionnement, pour faire face à l'expansion.

Même problème de visibilité financière à long terme que les autres consortiums, comment continuer à construire des choses à long terme.

<http://cahier.hypotheses.org/>

- Archives des ethnologues (Judith Hannoun)

Objectifs du consortium créé en 2011 : traiter les matériaux de terrain d'ethnologues et les valoriser, dans un souci d'interopérabilité ; s'assurer que les archives d'ethnologues sont préservées et valorisées dans les laboratoires et que les chercheurs peuvent y accéder ; créer une communauté autour des archives. Réfléchir à la place des bibliothèques. Collecter, sauvegarder, valoriser : donner une orientation pour suivre les standards, coordonner les pratiques.

Au début les différents partenaires du consortium (5 au départ) ne se connaissaient pas : il a fallu construire la communauté et développer des projets communs. Aujourd'hui il y a 8 partenaires sur toute la France.

Il y a environ 150 fonds d'archives à gérer (> 300 000 fichiers numériques) et beaucoup d'opérations ont été réalisées (6 plateformes de diffusion interopérables, + de 10 000 consultations, 1 colloque européen, publications, groupe Dariah, 1 site de transcription collaborative <http://transcrire.huma-num.fr>, numérisation importante, cartographie des terrains des ethnologues <http://ethnologues.huma-num.fr>, un réseau Éthique et droit <http://ethiquedroit.hypotheses.org> ).

La dimension éthique est centrale dans les objectifs du consortium. Plusieurs partenaires travaillent au sein du réseau Éthique et Droit qui traite des problématiques juridiques rencontrés lors de la diffusion et le traitement des données. L'objectif du réseau est de faire un guide des bonnes pratiques.

Les avancées sont la constitution d'une communauté, l'accès à de nouvelles sources (production de corpus numériques, sources inédites) qui ont eu comme effet la possibilité de nouveaux projets de recherche, il y a aussi eu un impact sur les politiques de recherche (ex: gestion des données de la recherche) et sur les financements (BSN5, ANR, etc.) grâce à l'effet consortium.

Le consortium des ethnologues est une sorte de laboratoire d'expériences, en mode projets, pas un mode routine. D'où des difficultés : internes sur le choix des méthodologies ; dans les relations avec HumaNum pour faire coïncider les objectifs que le consortium s'est donné avec les demandes d'HumaNum (livrables et productivité).

Mais ceci a permis la création d'une communauté identifiée.

Interrogation sur les publics en matière de réutilisation. Un apport en termes d'effets sur des possibilités de financements qui n'auraient pas été obtenues autrement.

<http://ethnologues.huma-num.fr/>

- ImaGEO (Caroline Abela)

Constat : il n'y avait pas, ou peu, de réutilisations des données cartographiques car elles étaient conservées dans des endroits dispersés (fonds des laboratoires de recherche, de bibliothèques) et sur des supports difficilement exploitables (ex: plaques de verre). Il s'agissait de données non géolocalisées, non indexées.

Mission : archivage de données géographiques anciennes cartographiques et géographiques qui sont un héritage du patrimoine historique de la discipline ; archiver des données contemporaines, pour les rendre accessibles, consultables et réutilisables (par ex. photo-comparaison pour voir l'évolution d'un paysage, la fonte d'un glacier, etc.).

Les membres du consortium sont des UMR et le SCD-carthothèque de Bordeaux Montaigne. Il y a une vingtaine de membres qui participent aux activités et un Comité de pilotage qui s'occupe des orientations scientifiques et techniques, du budget etc.

2 plateformes de diffusion : [1886](#) (bibliothèque numérique) et [AOMS](#) (cartothèque). Il y a aussi MediHal pour les photographies.

Principe : a minima rendre accessibles les métadonnées.

Le périmètre de l'objet de ce consortium sont les données cartographiques produites par les chercheurs des laboratoires (valeur patrimoniale).

Le consortium participe à un programme DARIAH.

Difficultés : peu de membres participants et charge de travail forte avec des ambitions importantes.

Solutions : recherche d'une expertise externe (séminaires, journées d'étude, etc.). En 2016 : création d'un outil de diffusion commun pour les cartes (réservoir commun de cartes ; a minima une interface de recherche spatiale commune aux cartes) ; problème des normes /schéma de métadonnées communes.

Autre question : relier les objets de recherche (il y a peu de chercheurs dans le consortium) à des préoccupations documentaires (il y a surtout des personnels de l'IST dans le réseau) et des publics oeuvrant dans ce sens.

Exemples de projets de recherche que le réseau peut favoriser : la photo-comparaison ; autre idée: des expositions dans les Pays du Sud où les cartes ont été réalisées à l'origine (par exemple à Madagascar).

4 ans, c'est court ; différents chantiers restent à réaliser comme autant de défis : valorisation, archivage pérenne, formation.

La plus grosse limite : la complexité de l'objet lui-même : il n'y a pas de normes de catalogage qui soient propres à l'objet "carte".

<https://imageo.hypotheses.org/>

- IRCOM

Consortium autour de la linguistique (il y en a deux) dont la mission est d'accompagner le développement et la valorisation de corpus oraux et multimodaux.

Le consortium est au terme de sa labellisation, en attente de la décision d'HumaNum concernant la nouvelle labellisation, avec fusion avec l'autre consortium de linguistique (corpus écrits). [depuis devenu Corpus, Langues et Interactions (CORLI)].

En 2012, au moment de la création du consortium, il y avait une communauté préexistante, plus organisée que dans le cas des autres consortiums, mais aussi des fonds identifiés. Par contre, il y avait une grande diversité des supports et un besoin de mise à disposition de ces fonds ainsi que de standards internationaux.

Le consortium est doté d'un Comité de pilotage (10 personnes, réunions mensuelles en téléconférence), d'un Comité scientifique (environ 30 personnes) et de groupes de travail. Il est apparu que les deux instances les plus efficaces ont été le CoPil et les groupes de travail, le Conseil scientifique étant peu actif et s'avérant redondant avec les groupes de travail. Organisation tournante de l'animation du consortium. Les groupes de travail (5 + 1) :

- > Finalités scientifiques
- > Interopérabilité
- > Multilingues
- > Multimodalité
- > Juridique
- > Dépôt et conservation

Comment fonctionnent les groupes de travail ? Les gens s'impliquent quand ils ont un vrai intérêt à s'impliquer, à faire bouger les choses.

Le consortium a permis de recenser environ 80 projets / corpus identifiés, de taille et de date très différentes. Il a aussi aidé à la finalisation de 30 projets (jusqu'au dépôt), auprès de centres de ressources.

Difficultés :

- > CS peu actif
- > mobilisation décentralisée et à distance des membres (solution : coupler chaque AG avec l'organisation d'ateliers pratiques ; faire tourner les membres du CoPil : tous les deux ans)
- > difficulté générale à mobiliser (peu de chercheurs dans le réseau)
- > problèmes des relations avec d'autres institutions et les autres opérateurs (COCOON, BNF, ORTOLANG) & de la distinction entre le rôle du consortium et celui des opérateurs
- > pérennisation

<http://ircom.huma-num.fr/site/accueil.php>

- MUSICA (Joséphine Simonnot & Aude Da Cruz Lima)

Ont démarré en juin 2014. Ont mis une année à monter le réseau. Pas de base disciplinaire mais base thématique (donc multidisciplinaire).

Partenaires : LESC (ethno-musicologues, données sonores et audiovisuelles, notation non-européenne), CESR (partitions de musique de la renaissance), LAM (physiciens, acousticiens de la musique), IRCAM (traitement du signal, annotation du son), IREMUS (musicologie et iconographie musicale), MMSH

Il y a 3 axes de travail :

- encodage des partitions / musique notée (MEI)
- Archivage, conservation et annotation des données sonores et vidéos (musique enregistrée)
- Classification et référentiel pour décrire les instruments de musique et la voix

Des équipes travaillaient sur ces thèmes avant la labellisation ; enjeu : dépasser les obstacles des différences disciplinaires pour faire travailler les gens ensemble.

Bureau = 1 représentant de chacun des axes. Organisation = 1 journée d'AG par axe ; de nouveaux projets peuvent proposer des adhésions.

Le consortium mélange, dans sa composition, des projets et des institutions.

Problématiques juridiques, récente journée thématique organisée sur ce thème, projet en lien avec la bibliothèque numérique Europeana.

Dimension interconsortium : le consortium Musica est déjà en lien avec IRCOM, Archives des ethnologues et Archives de mondes contemporains.

Il a le projet de réaliser des ateliers de formation pour la vidéo.

Les difficultés rencontrées : solliciter des personnes qui sont déjà très occupées, faire communiquer chercheurs et professionnels de l'information ; savoir comment le CNRS se positionne sur la question de la propriété intellectuelle des données de la recherche.

<http://humanum.hypotheses.org/503>

- 3D (Mehdi Chayani)

Labellisé en 2014.

L'usage de la 3D dans les SHS est né à la fin des années 80 avec un programme de recherche CNRS qui portait sur une étude des temples de Karnak et qui a permis de retranscrire en 3 dimensions des informations collectées lors des campagnes de fouilles archéologiques. Cela a fait naître de nouvelles pistes méthodologiques pour les chercheurs en SHS.

Composé de 10 équipes (notamment des UMR et une UMS Archéovision, l'INRAP) ; couverture assez large au niveau national ; diversité des disciplines et des objets de recherche portés par les



partenaires (archéologues, architectes, historiens, ingénieurs 3D, informaticiens), grande diversité des profils et des usages de la 3D, ou de l'expérience / expertise sur ce plan.

Le champ d'action du consortium touche surtout l'archéologie au sens très large. Mais d'autres disciplines qui ne sont pas couvertes pas le champ du patrimoine et qui relèvent des SHS peuvent très bien intégrer le consortium. Reste à les identifier.

Le consortium est coordonné par Archéovision et fonctionne avec un bureau décisionnaire au sein duquel tous les membres sont représentés, il y a également un coordinateur.

Missions du consortium :

- > Sensibiliser les acteurs de la recherche SHS aux enjeux liés aux usages des modèles 3D,
- > Clarifier le vocabulaire associé aux technologies 3D pour les SHS
- > Suivre les usages et la production scientifiques autour de la 3D en SHS
- > Mutualisation des ressources
- > Développer des outils spécifiques 3D SHS en open source
- > Guides de bonnes pratiques (méthodologie, formats, schéma métadonnées, etc.)
- > Favoriser les partenariats pluridisciplinaires (informatique/SHS).

Livrables :

- > Mise en place de procédures de dépôt de données 3D SHS
- > Mise en place d'un conservatoire 3D
- > Guides de bonnes pratiques
- > Mise en place de cahiers des charges pour les projets 3D
- > Identification des contextes scientifiques des productions 3D et lieu de dépôt des modèles 3D.

Difficultés :

- > Large spectre de profils avec des expériences et des compétences très variées
- > Difficulté dans la mise en place de mutualisation de certains logiciels 3D.

<http://humanum.hypotheses.org/480>

### ***Premiers échanges et préparation des ateliers de l'après midi***

- Thomas Labarbé : question de la mise en relation de nos consortiums entre eux mais aussi avec les infrastructures européennes... Par exemple Darjah.

Ces questions institutionnelles pourraient faire l'objet d'un des ateliers de l'après-midi.

- C. Abela : implication d'Imagéo s'est faite suite à une présentation de la Dariah par Huma-Num et des relations entre différents consortiums. Cela a abouti au projet de rédaction d'un guide de bonnes pratiques.

- O. Baude : les consortiums sont au coeur de la TGIR. La TGIR a pour mission d'accompagner les consortiums. Encourage les activités interconsortium. Les consortiums n'ont pas vocation à être des structures pérennes, c'est pour cela que le Conseil Scientifique d'Huma-Num oriente les activités des consortiums vers la production de livrables. Ne veut pas limiter l'évaluation à la seule mise en réseau (même si reconnaît que les moyens alloués sont relativement faibles).

- S. Pouyllau : matinée positive car on voit qu'il y a une vraie dynamique qui arrive à s'ancrer dans les communautés scientifiques.

Mise en place des rencontres Huma-Num du 10 au 13 octobre 2016 (3 jours et demi) ; tous les consortiums sont invités, ainsi que d'autres partenaires (MSH, etc.).

L'ensemble des points discutés pendant la présente journée pourront être approfondis durant ces journées. Il sera aussi question de l'implication dans les projets européens comme Dariah & Parthenos.

- Questions des ateliers de l'après-midi
  - Juridique et éthique
  - Référentiel communs et interopérabilité
  - Archivage pérenne
  - Portails de publication et valorisation

## **Ateliers**

### 1. Problèmes éthiques et juridiques

Plusieurs consortiums étaient représentés : Musica, IRCOM, Cahier, archiPolis, ArcMC, archives des ethnologues.

Nécessité de mutualiser les guides de bonnes pratiques, documents, types de contractualisation, conventions de dépôt ou de réutilisation, formulaires de consentement, etc. produits par les uns et les autres.

Il apparaît toutefois illusoire d'aboutir à une sorte de référentiel transversal à l'ensemble des consortiums, et donc des disciplines impliquées ; en revanche un socle commun peut être dégagé, partagé, à chacun ensuite d'adapter en fonction des spécificités de son domaine.

Discussion sur la possibilité de dégager des "pôles" au sein de l'ensemble des consortiums d'Humanum eu égard à ces questions, fondés sur des affinités plus ou moins fortes.

Discussion sur la nécessité de partir d'interrogations concrètes qui intéressent pratiquement les chercheurs, les acteurs impliqués par ces questions, et non des "guides" descriptifs : formule du jeu

“questions / réponses”. Intérêt de trouver des témoignages, des retours d’expérience de chercheurs, sur la question de l’usage des formulaires de consentement en SHS.

Proposition que ce soient les consortiums qui financent la rédaction d’une sorte de “**guide**” **commun**, par des juristes spécialistes. Il faudrait organiser pour ce faire un atelier de travail transversal aux consortiums.

Autre stratégie d’action discutée : s’unir pour “pousser le droit”, rencontrer les instances décisionnaires (CNRS, CNIL..) et **faire pression** sur elles pour justifier la mise en place de régimes juridiques plus favorables à la recherche en SHS, notamment en ce qui concerne les questions de propriété intellectuelle, d’anonymisation / de confidentialité des données archivées en vue d’une réutilisation à court terme. Idée que la France est le pays qui, en Europe, a un des régimes les plus stricts en matière de propriété intellectuelle et de protection des données personnelles.

## 2. Référentiels communs et interopérabilité

Quels sont les référentiels externes utilisés ? Quels sont les référentiels élaborés en interne et potentiellement réutilisables par d’autres ? Quels sont les besoins de référentiels non satisfaits ? Quelles sont les connaissances et compétences en matière d’utilisation et de conception de référentiels ? Les besoins de formation et d’information ?

Protocole d’échange OAI-PMH (choix des jeux de métadonnées).

Exposition des métadonnées sur le modèle du web de données.

L’atelier a permis de constater qu’il y avait des niveaux de maîtrise du domaine très différents d’une communauté à l’autre. Les **bibliothécaires et documentalistes** sont les plus avancés dans le domaine, disposant d’outils communs à la profession et très puissants. Les autres communautés ont des connaissances plus superficielles et n’ont pas accès aux outils des documentalistes. De plus les conditions de description de données peuvent être très différentes d’un usage et d’un type de support à l’autre.

La conclusion de l’atelier est qu’il y aurait un double besoin :

- organiser une mission d’expertise auprès des différents domaines et acteurs pour présenter l’existant et faire connaître les outils et normes disponibles quand ils existent,
- permettre aux spécialistes que sont les documentalistes de présenter aux autres communautés leur manière de travailler et leurs outils.

## 3. Archivage pérenne

Qui peut et comment utiliser les services d’Huma-Num (opérateur CINES) pour la conservation intermédiaire ou définitive de données numériques ?

Comment faire pour des fonds mixtes (supports traditionnels et données numériques) ?

Comment sont organisés chez chacun, les fonds ? Fonds de personnes (chercheurs), Fonds thématiques (projets, secteurs), Fonds organiques (laboratoires)...

Quelle communicabilité pour les données (archives publiques, archives privées...) ?

Comment organiser le rôle de relais des besoins des communautés scientifiques auprès d'Huma-Num ?

La définition de l'archivage pérenne si on se réfère à la théorie des 3 âges des archives c'est aller vers l'âge définitif. N'y vont que les archives qui ont un **intérêt** à être conservées définitivement et c'est une mission qui est confiée au réseau des archives publiques, si le producteur ne peut pas s'en charger c'est le réseau des archives publiques qui s'en occupe.

Le CINES n'a **pas** vocation à conserver les documents sans limite de temps ; au bout d'un moment ils ont vocation à être versés aux Archives de France.

Le discours du CINES est qu'on ne sait pas conserver les supports numériques, il faut donc se focaliser sur la préservation de **l'information** inscrite sur ces supports.

Donc est-ce qu'il faut dire aux chercheurs de conserver leur ancien support => non pas forcément il vaut mieux mettre l'information à plusieurs endroits.

Est-ce que les chercheurs doivent conserver leurs originaux ?

Il apparaît que la question de l'archivage pérenne, si on creuse, est moins claire qu'on ne pourrait le penser. On doit réfléchir non pas seulement à la sauvegarde mais à l'**accès** (moissonnage, etc. ). Si on ne fonctionne pas en réseau tout l'édifice cesse de fonctionner.

Quand on parle d'archives pérennes, qu'est-ce qu'on conserve ?

Est-ce que le fait de parler d'archivage pérenne, ce n'est pas aussi de conserver les outils qui permettent de lire les données ?

C'est le lien avec les utilisateurs qui semble primordial (c'est bien de pouvoir encore ouvrir des fichiers mais on ne peut pas toujours les exploiter dans des logiciels : du coup personne ne s'en sert). Les chercheurs ne sont pas intéressés par l'archive pour l'archive. Il faut qu'ils soient associés à l'idée de valorisation de leur travail, de réutilisation des archives etc.

L'archivage pérenne ce n'est pas seulement garantir leur accessibilité mais aussi leur **intelligibilité**. C'est énormément de travail pour la constitution de paquets d'informations qui vont être traités dans des chaînes de pérennisation. On a très peu de moyen et c'est une activité peu valorisée dans les carrières des chercheurs. Est-ce qu'il y a des choses qu'on peut faire à minima pour faire ce travail ? La question est aussi la valorisation de ce travail pour le chercheur. Il va falloir faire un peu de lobbying pour faire valoir ce travail. De plus, il faudrait développer des formations pour que les chercheurs arrivent avec leur documentation.

Il ressort finalement deux questions :

- jusqu'où est-on déjà engagés dans des projets d'archivage pérenne ?
- quels besoins faire remonter à Huma-Num ?

Le consortium Archives des ethnologues (regroupement de plusieurs bibliothèques partenaires conservant des fonds d'archives de chercheurs) a réfléchi à l'archivage à long terme des fichiers images de ses partenaires (produits suite à opérations de numérisation) via la TGIR et le CINES.

Il y a d'abord la question des moyens en personnel pour la TGIR afin d'accompagner les producteurs dans le versement des fichiers à préserver (la personne en charge des projets de dépôt au Cines est surchargée).

Il leur est aussi apparu que, pour chaque partenaire, l'une des difficultés principales n'est pas tant la préparation des SIP (paquet d'information soumis à l'archive) ou la mise en place d'un plan de classement (auquel il faut bien réfléchir), mais la question de la gestion de la soumission de ces paquets, du contrôle du workflow (suivi des dépôts, historique des versements, mises à jour des paquets existants, corrections/ajouts de métadonnées, etc). Ils ne disposent pas, dans leurs systèmes d'information, d'outils ou de modules qui permettraient de gérer les dépôts au Cines et d'en garder trace, ni même d'outils permettant d'automatiser la création des SIP et les opérations nécessaires aux versements (création d'empreinte de fichier par exemple).

Il y aurait ainsi une demande possible auprès d'Huma-Num d'un outil pour gérer les collections numériques de plusieurs acteurs et qui intégrerait des modules utiles à la préparation et à la gestion des SIP et des versements.

Concernant l'organisation/plan de classement du consortium Archives des ethnologues au CINES : il est pensé en rapport (miroir) avec les inventaires EAD des fonds d'archives traités en amont (les originaux). Il y aurait 1 projet (ppdi) par bibliothèque avec, très schématiquement, une structure hiérarchique : un SIP "parent" (pour 1 fonds d'archives de chercheur) vide et qui serait relié (possibilité d'indiquer des relations entre les SIP) à des SIP "enfants", correspondant à des "levels" des inventaires EAD, dans lesquels fichiers images et métadonnées métiers seraient insérés.

Il est également question de l'entrepôt intermédiaire qui précède l'archivage pérenne.

Le consortium Cahier a, d'abord, archivé les métadonnées.

Le consortium IRCOM a été pilote pour l'archivage pérenne en faisant avec Huma-Num une étude sur les formats audio.

- La question qui reste est "de quelle manière mutualiser" la réflexion sur cette problématique de l'archivage à long terme ?

- Faire un panorama des opérateurs ? Importance de repérer les opérateurs à même de s'occuper des versements au Cines, voir les manques.

- Plan de gestion de données numériques : réflexion sur les bonnes pratiques à faire passer auprès des producteurs et partenaires des consortiums.

Est-ce que les consortiums ne pourraient pas fournir des éléments qui remontent des communautés scientifiques sur les besoins thématiques.

- Les consortiums ont un rôle de relais entre les communautés scientifiques et les opérateurs de l'archivage pérenne (dont la TGIR).
- Action de formation.

#### 4. Portails de publication et valorisation

Impératif : donner de la visibilité à ce qui est réalisé par les consortiums.  
Distinguer les différents publics cibles auprès de qui valoriser :

- **les consortiums eux-mêmes** : aller vers une plate-forme d'échanges, une sorte de forum des consortiums ? Permettrait d'inscrire dans la durée les ateliers par exemple et d'encourager les échanges d'expertise.
- **communauté de la recherche au sens large**, en France et ailleurs. L'objectif est de donner une visibilité aux outils et plateformes créés par tous, et ce pour plusieurs raisons :
  - augmenter l'usage de ces outils à des fins de recherche
  - en faire des outils pédagogiques, en lien avec les UFR, Écoles Doctorales, etc., voire les lycées ; notamment en axant sur les enjeux épistémologiques
  - associer de nouveaux laboratoires et chercheurs
  - encourager la documentation et l'archivage de la recherche, notamment auprès des publics de l'IST.

Des chercheurs ont exprimé un manque de clarté du site web d'Huma-Num présentant les consortiums. Suggestion de modifier la présentation actuelle en utilisant par exemple un tableau qui récapitule de manière plus synthétique l'objet / le périmètre / le type de données / les livrables... de chaque consortium. Autre suggestion : mettre en ligne sur le site d'Huma-Num des slides (remaniées) des présentations de la matinée.

- **grand public** : peut être associé à l'activité de certains consortiums, notamment pour encourager le crowdsourcing des documents signalés (ex : CAHIER travaille sur ces questions). L'intérêt est cependant plus limité et moins prioritaire d'un point de vue collectif.

Améliorer la visibilité et la communication, notamment en s'appuyant sur des structures pérennes comme les MSH ou les SCD, permettrait de donner une place plus importante aux consortiums, sur le plan local mais aussi national, Il faut communiquer sur la plus-value apportée par les consortiums dans la gestion des données, en faire un argument pour l'obtention d'ANR (exemple : Archipolis).

Quid de l'identification des opérateurs privilégiés des consortiums autour de la mise à disposition, des usages des données traitées par les consortiums ? A 2, 3, 4 ans ? (MSH, Equipex, etc.); cela doit être pensé par la TGIR.

Ces enjeux sont aussi liés aux questions de l'accès unifié, de la normalisation et harmonisation des instruments de recherche (dont indexation), de l'interopérabilité, traitées dans d'autres ateliers.

Actions envisagées :

- réalisation d'une cartographie des consortiums à diffuser sur le site Huma-Num
- outil d'échanges de pratiques inter-consortium (forum, listes de diffusion), pour identifier qui fait quoi, qui a quelles compétences, etc...

### ***Synthèse des ateliers et perspective de la poursuite des échanges entre consortiums***

Quelles suites donner à cette journée ?

Est-ce que la journée annuelle d'Huma-Num recouvre les problématiques de cette journée ?

O. Baude : oui

Est-ce qu'on peut imaginer une suite notamment aux ateliers qui ont fait émerger des pistes de travail ?

S. Wolikow: proposition de faire une valorisation des CR des ateliers plus large que les petites synthèses de la fin.

Réalisation transversale par petits groupes (impossible à 11) puis faire un événement ou quelque chose dans 6 mois autour de ces travaux transversaux. Il faudrait faire participer Huma-Num non pas seulement en tant que TGIR tutelle. Les consortiums les plus avancés dans tel ou tel domaine pourraient être moteur et organiser des après-midi indépendants etc.

Il serait effectivement intéressant d'avoir un interlocuteur à Huma-Num qui participe avec nous à ces réflexions

S. Wolikow : question des plateformes intermédiaires qui doivent pouvoir faire face à nos demandes. Plaide aussi pour la coopération de certains consortiums sur les problématiques de la revisite et de la réutilisation des archives de la recherche

23 juin il y a une journée d'études à Montpellier sur le thème "Archiver la recherche : responsabilité partagée". Il serait intéressant de se retrouver après cette journée qui est aussi le moyen de se rencontrer. La question est alors de savoir comment se retrouver, et sous quelle forme organiser la rencontre.

- **Conclusion (B. Denni et S. Wolikow)**

Question de la formation au sein des écoles doctorales : les consortiums pourraient intervenir à cet endroit

On peut déjà mettre en place une plate-forme pour échanger, mais il faut aussi qu'on se rencontre (durant les journées de juin ?).

Comment mettre en commun nos procédures et notre gestion des données et des archives sans gommer les diversités disciplinaires ?

Préparer les journées d'automne d'HN 10-13 octobre.