



Huma-Num. La TGIR des humanités numériques. Rapport d'activité 2013-2015

Marc Renneville, Stéphane Pouyllau

► To cite this version:

Marc Renneville, Stéphane Pouyllau. Huma-Num. La TGIR des humanités numériques. Rapport d'activité 2013-2015. [Rapport Technique] TGIR Huma-Num (UMS 3598). 2015. <halshs-01390938>

HAL Id: halshs-01390938

<https://halshs.archives-ouvertes.fr/halshs-01390938>

Submitted on 2 Nov 2016

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.



#1

Huma-Num est une très grande infrastructure (TGR) née de la fusion du TGE-Adonis et de IR-Corpus. Elle vise à faciliter le tournant numérique de la recherche en sciences humaines et sociales.

Pour remplir cette mission, la TGR Huma-Num est bâtie sur une organisation originale consistant à mettre en œuvre un dispositif humain (concertation collective) et technologique (services numériques pérennes) à l'échelle nationale et européenne en s'appuyant sur un important réseau de partenaires et d'opérateurs.

En savoir plus



TGIR Huma-Num
Pôle communication
190, avenue de France
75013 PARIS

huma-num.fr

INFORMER

PARTAGER

DIFFUSER



SOMMAIRE

Contexte d'activité de la TGIR Huma-Num (2013-2015)	5
I - Mission de la TGIR Huma-Num	7
II - Organisation de l'infrastructure	7
III - Répartition des services numériques.....	12
IV - Dimension internationale	16
V - Perspectives	19
VI - Liste des consortiums au 30 juin 2015.....	21

Direction d'Huma-Num

Marc Renneville,
Directeur de recherche CNRS,
directeur.

Stéphane Pouyllau
Ingénieur de recherche CNRS,
directeur-adj. technique .

Contexte d'activité de la TGIR Huma-Num (2013-2015)

Prévue dans la feuille de route 2008-2012 du MESR sur les infrastructures de recherche, la décision de fusionner le TGE Adonis (créé en 2004) et l'IR Corpus (créé en 2011) reposait sur le constat de la complémentarité des services et des réalisations des infrastructures précitées : par la création et la réflexion sur les corpus, les consortiums d'IR-Corpus tendaient à favoriser l'acculturation du numérique dans les communautés SHS. Mené par les chercheurs, le processus de concertation faisait émerger des besoins en partie couverts par les outils précédemment forgés par le TGE Adonis : grille de services et de traitement de données, archivage à long terme, portail d'accès unifié aux données SHS, mise en réseau au niveau européen. La nouvelle TGIR devait donc mettre en synergie la concertation collective des communautés scientifiques et l'offre d'outils numériques.

La réussite de l'opération de fusion pouvait ainsi être considérée comme acquise dès lors que la nouvelle TGIR aurait atteint les trois objectifs suivants :

- Objectif 1 : la reprise et la consolidation des missions et des activités opérées par le TGE Adonis et Corpus IR.
- Objectif 2 : le développement de nouveaux services dédiés aux données de la recherche.
- Objectif 3 : offrir rapidement une bonne visibilité de la nouvelle infrastructure dans les communautés scientifiques SHS.

La fusion a été réalisée au 1er mars 2013. Deux années après, ce bilan intermédiaire a pour but de mesurer le chemin accompli et d'apprécier la dynamique engagée par la TGIR Huma-Num.

I - Mission de la TGIR Huma-Num

Huma-Num est une très grande infrastructure (TGIR) visant à faciliter le tournant numérique de la recherche en sciences humaines et sociales.

Pour remplir cette mission, la TGIR Huma-Num est bâtie sur une organisation originale consistant à mettre en œuvre un dispositif humain (concertation collective) et technologique (services numériques pérennes) à l'échelle nationale et européenne en s'appuyant sur un important réseau de partenaires et d'opérateurs.

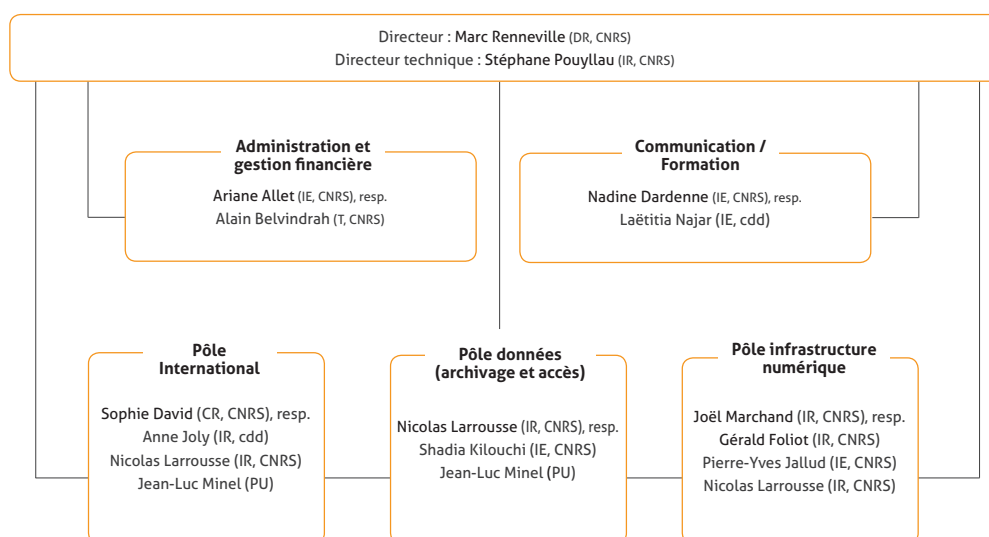
La TGIR Huma-Num favorise ainsi, par l'intermédiaire de **consortiums** regroupant des acteurs des communautés scientifiques, la coordination de la production raisonnée et collective de corpus de données (recommandations scientifiques, bonnes pratiques technologiques). Elle développe également un dispositif technologique unique permettant le traitement, la conservation, l'accès et l'interopérabilité des données de la recherche en SHS. Ce dispositif est composé d'une grille de services, d'une plateforme d'accès unifié (**ISIDORE**) et d'une procédure d'**archivage à long terme**, dédiés aux besoins des communautés SHS.

La TGIR Huma-Num propose en outre des **guides de bonnes pratiques** technologiques généra-listes à destination des chercheurs. Elle mène régulièrement des actions d'expertise et de **formation**. Elle porte la participation de la France dans le projet DARIAH en coordonnant la contribution nationale.

II - Organisation de l'infrastructure

Statut de la structure : UMS 3598 (tutelles CNRS-Aix-Marseille Université-EPCS Condorcet) suite à la fusion de l'UPS 2916 (ADONIS) et de l'UMS 3534 (CORPUS) au 01/03/2013.

1. Organigramme



2. Gouvernance

La TGIR Huma-Num est dotée d'un comité de pilotage et d'un conseil scientifique.

Comité de pilotage

Missions :

- approuver les programmes scientifiques ;
- valider la proposition de budget annuel dans le cadre d'une programmation pluriannuelle et de son exécution ;
- mettre en place un conseil scientifique et un comité des utilisateurs ;
- mettre en place les instances d'évaluation requises ;
- approuver les projets de collaboration ;
- préparer l'évolution et l'eupéanisation de la TGIR Huma-Num

Il comprend neuf membres :

- Patrice Bellot, Représentant de Monsieur le Président de l'Université Aix-Marseille
- Olivier Baude, Président du Conseil Scientifique
- Patrice Bourdelais, Directeur de l'Institut des Sciences Humaines et sociales (InSHS)
- Christian Chardonnet, Chef du département Grandes Infrastructures de Recherche, Ministère de l'Enseignement Supérieur et de la Recherche
- Alain Fuchs, Président de l'Alliance Athéna
- Bertrand Jouve, Directeur adjoint scientifique, en charge des TGIR, Institut des Sciences Humaines et sociales (InSHS)
- Sylvie Max-Colinart, adjointe au Chef de la Mission, Ministère de la culture et de la communication/SG/SCPCI/Département de la recherche, de l'enseignement supérieur et de la technologie
- Laurent Pinon, Représentant du Président du comité des très grands équipements scientifiques et grandes infrastructures (CNRS)
- Jean-Claude Waquet, Président du Campus Condorcet

Mokrane Bouzeghoub, représentant de l'INS2I (CNRS) et la direction de la TGIR Huma-num assistent aux réunions du Comité de pilotage.

Conseil scientifique

Le conseil scientifique donne son avis sur toute question scientifique et technologique concernant les activités de la TGIR Huma-Num.

En particulier, le conseil scientifique :

- donne un avis d'ensemble sur les priorités de la TGIR Huma-Num ;
- évalue les propositions d'actions soumises par la direction de la TGIR Huma-Num ainsi que les procédures retenues (type d'appels d'offres, évaluation, etc.) ;
- évalue annuellement l'activité de la TGIR Huma-Num et précise les critères et les indicateurs qui lui paraissent pertinents ;
- peut procéder, à la demande du comité de pilotage, à des évaluations ponctuelles.

Il comprend douze membres :

- Olivier Baude, Maître de conférences, Laboratoire Ligérien de Linguistique
- Frédérique Brunet, Chargée de recherche au CNRS, ArScAn, équipe "Archéologie de l'Asie Centrale"
- Philippe Casella, Directeur de la direction du développement de la recherche, EHESS
- David Fiala, Maître de conférences, Centre d'Etudes Supérieures de la Renaissance
- Françoise Genova, Directrice de recherche au CNRS, Observatoire astronomique de Strasbourg
- Brigitte Marin, Directrice de la Maison Méditerranéenne des Sciences de l'Homme
- Frédéric-David Martin, Conservateur d'Etat des bibliothèques, Bibliothèque nationale de France
- Pierre Mounier, Directeur-adjoint, Centre pour l'Édition Electronique Ouverte
- Yann Nicolas, Chef du département Etudes et Projets, Agence Bibliographique de l'Enseignement Supérieur
- Olivier Rouchon, Responsable du département Archivage et diffusion, Centre informatique nationale de l'enseignement supérieur
- Julien Sicot, Responsable du système d'information documentaire, Université de Rennes 2
- Marcello Vitali Rosati, professeur au département des littératures de langue française à l'université de Montréal

Le conseil scientifique est présidé par Olivier Baude. Françoise Genova en est la vice-présidente.

La direction de la TGIR Huma-Num participe aux réunions du conseil scientifique. Alexandre Gefen, représentant du conseil scientifique de l'INSHS est invité permanent de ce conseil.

Site web : <http://www.huma-num.fr/>

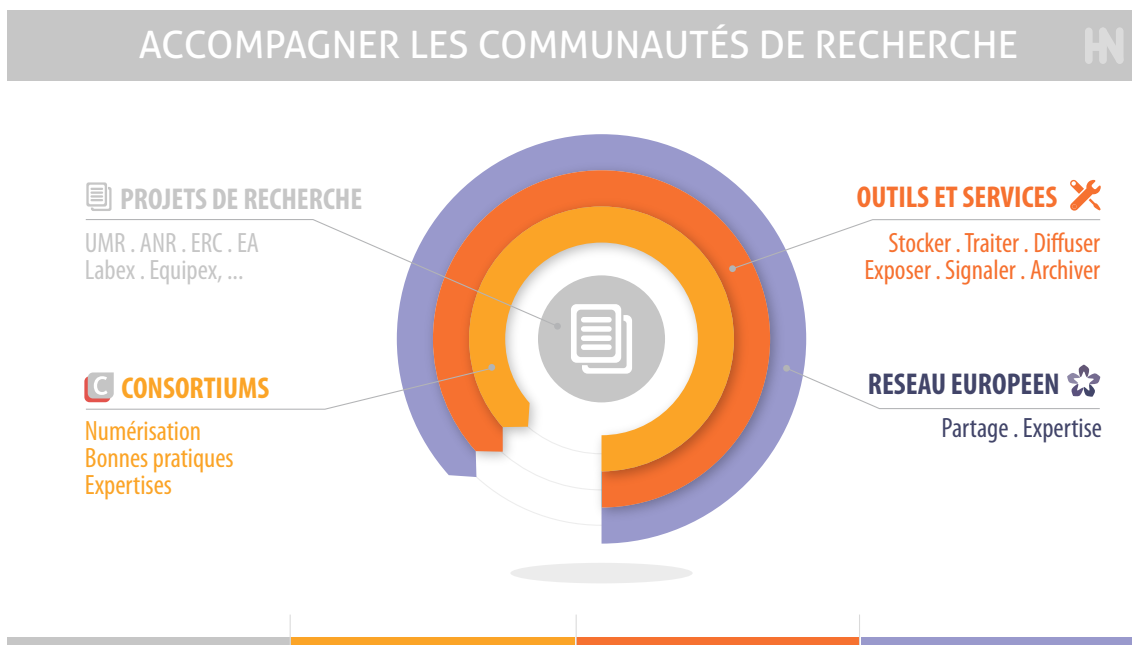
3. Contexte local, français et international

La TGIR Huma-Num travaille au niveau national avec :

- le CC-IN2P3 (Centre de Calcul de l'IN2P3) pour l'exploitation de l'infrastructure numérique
- le CCSD (Centre pour la Communication Scientifique Directe) pour la mise en œuvre de la plateforme de recherche ISIDORE
- le CINES (Centre Informatique National de l'Enseignement Supérieur) pour l'archivage à long terme
- L'ABES (Agence Bibliographique de l'Enseignement Supérieur) pour le travail sur l'exposition et le signalement des données (Star, Numes, Sudoc...)
- Le réseau des Maisons des Sciences de l'Homme pour l'organisation des consortiums
- Les plateformes de publications scientifiques Cléo/revues.org et Persée.
- Le CLEO (Centre pour l'édition électronique ouverte)

Les activités du pôle international d'Huma-Num s'inscrivent dans le cadre de politiques scientifiques établies aussi bien au niveau européen qu'au niveau national.

4. Concertation collective



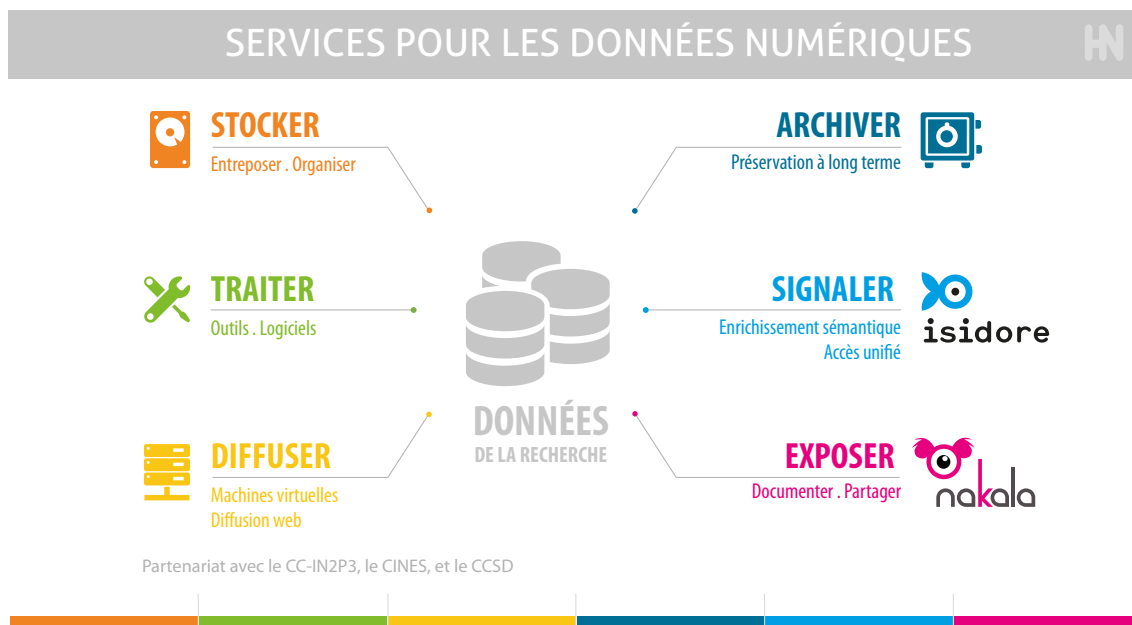
- La TGIR est centrée sur la mise à disposition de services numériques permettant l'organisation, l'exploitation et la valorisation de la production des données numériques de la recherche en SHS. C'est à ce titre qu'elle contribue à l'organisation de la concertation des communautés scientifiques sur les enjeux du numérique.

La concertation collective organisée par la TGIR s'appuie sur 11 consortiums disciplinaires (voir la liste en fin de document) regroupant plus 120 équipes (UMR, UMS, équipes d'accueil universitaires, Maison des sciences de l'Homme) dotant les communautés scientifiques des SHS d'un réseau d'échanges et de travail sur les questions autour de l'appropriation du numérique pour les données de la recherche et leurs outils de traitements (corpus ; signalement ; diffusion). L'ensemble des activités des consortiums nourrit et oriente la stratégie de déploiement de services de la TGIR (cf. ci-dessous) dans le but de les rendre accessibles au plus grand nombre et de les ancrer dans une forte légitimité scientifique. Les consortiums sont labélisés sur un programme d'activité et pour une durée de 4 ans par le comité de pilotage, après évaluation du conseil scientifique. L'activité 2014 des consortiums est détaillée dans les rapports des consortiums (ils peuvent être obtenus sur demande à la TGIR) et au travers des évaluations annuelles faite par le conseil scientifique (cf. rapport du président du conseil scientifique en décembre 2014). En 2015, 4 consortiums – lancés en 2011 - finissent leur période de labellisation ; un processus d'évaluation des résultats a été mis en place et sera effectué par le conseil scientifique de la TGIR.

Les consortiums et plus généralement les communautés ou équipes scientifiques réalisant des projets de recherche, peuvent s'appuyer sur un ensemble de services pour le traitement, le stockage, la diffusion et la conservation des données numériques de la recherche en sciences humaines et sociales. La seconde vocation de la TGIR Huma-Num est de développer, de gérer et de maintenir en phase avec les besoins des communautés SHS un ensemble de services couvrant l'intégralité des besoins liés au cycle de vie des données de la recherche (voir ci-dessous).

III - Répartition des services numériques

La TGIR déploie un ensemble de services numériques ouverts, accessibles à l'ensemble des communautés scientifique SHS de l'ESR. Les services sont organisés autour de 6 grandes thématiques et regroupent chacun d'entre-eux, un ou plusieurs outils.



Stocker : Il s'agit de solutions pour entreposer et organiser les données numériques dans un espace sécurisé professionnel. Ce service est réalisé en partenariat avec le centre de calcul In2p3/CNRS.

Traiter : Ensemble de solutions adaptées aux besoins de transformations ou d'analyses sur les données en SHS : logiciels de traitement, de visualisation (calcul de réseaux, tuilage de données), d'encodage, de projections cartographiques, d'accès à de la puissance de calcul, etc.

Diffuser : Il s'agit ici d'une large palette de solutions pour diffuser les données. L'offre vise à couvrir tous les niveaux de besoins, du plus basique au niveau expert. Elle va de « packs » logiciels préconfigurés et optimisés pour l'écosystème des SHS : bibliothèques numériques, interopérabilité Europeana ; Gallica ; Isidore, etc) jusqu'à la fourniture de machines virtuelles préfigurées pour les communautés SHS (exemple : mise en œuvre de BaseX pour la diffusion de corpus XML/TEI). Ce service est interopérable avec les services d'exposition de données de la TGIR (NAKALA) et de signalement (ISIDORE).

Archiver : Huma-Num fournit conseil et accompagnement dans le processus d'archivage à long terme des données numériques : aide aux choix des formats de données, normalisation des métadonnées, transfert dans le système d'archivage du

CINES. Ce service est opéré dans le cadre d'une convention, renouvelée en 2014 pour 4 ans, entre la TGIR Huma-Num, le CINES et le SIAF (Service Interministériel des Archives de France). L'expertise de la TGIR Huma-Num est aujourd'hui largement reconnue sur ces aspects d'archivage des données numériques, au-delà même des SHS. En 2015, l'archivage de plusieurs programmes nationaux est en cours (OpenEdition, IRHT, Equipex Ortolang) ainsi que l'archivage des données d'École Française à l'étranger (EFEO). Pour rappel, c'est à l'aide de ce dispositif que les données de l'archive ouverte nationale HAL sont archivées.

Signaler : Créé en 2010, ISIDORE est un service multilingue de référence pour le signalement, l'enrichissement, et l'accès aux données en libre accès de la recherche en sciences humaines et sociales. Depuis janvier 2015, ISIDORE est devenu une plateforme internationale avec une nouvelle version permettant de collecter, enrichir, indexer les données numériques des sciences humaines et sociales en trois langues : anglais, espagnol, français. Cette fonction permet à un chercheur non-francophone d'avoir accès à des enrichissements et des classifications en anglais et en espagnol lui permettant de mesurer l'intérêt de ressources en langue française qu'il est en train de rechercher. Elle fait d'ISIDORE un moteur de recherche scientifique innovant et reconnu sur le plan international. En effet, motivés par le fait de relier et de signaler leurs données et documents dans un environnement de plus trois millions de ressources en SHS, plusieurs partenaires européens et internationaux ont souhaité participer et sont intéressés par les enrichissements trilingues reliant les données entre elles mais également disponibles pour créer des applications dans le web sémantique. Le CSIC (Madrid, Espagne)¹ était déjà présent dans ISIDORE depuis le mois de février et depuis la plateforme de livres en libre accès OAPEN (Amsterdam, Pays-Bas) est présente dans ISIDORE et y expose plus de 2000 ouvrages issus des disciplines SHS. Plus récemment², la plateforme SOAS (*University of London*, UK), les archives royales d'Espagne (Madrid, Espagne), les archives institutionnelles de l'Université de València – la plateforme RODERIC – et la bibliothèque numérique de *Bibliotecas de Castilla y León* (Espagne) ont également rejoint ISIDORE.

Par ailleurs, deux nouvelles fonctionnalités ont été développées pour la version pour tablettes et mobiles dans le cadre d'expérimentations sur la visualisation des données issues des enrichissements trilingues : les principales disciplines associées à un auteur sont maintenant représentées sous forme d'un radar disciplinaire. Cette visualisation graphique permet de se faire une idée rapide du profil d'un chercheur dans Isidore. Depuis mai 2015, un module - Isidore Suggestions - a été développé pour le système de gestion de contenu web (CMS) WordPress (le plus utilisé au monde). Ce module permet d'afficher sur la page d'un article des suggestions de ressources enrichies venant d'ISIDORE. Il est possible de cibler la recherche en sélectionnant une ou plusieurs disciplines scientifiques des SHS. Depuis sa publication sur WordPress.org, la plateforme internationale des modules pour ce CMS en mai 2015, ce module a été installé plus de 100 fois et plusieurs mises à jour ont eu lieu afin de répondre aux demandes des premiers utilisateurs. Des discussions sont en cours avec OpenEdition afin de proposer ce module dans la plateforme de carnet de recherche hypotheses.org.

¹ Voir <http://tinyurl.com/pw75ou3> [d.c. 2015/06/12].

² Voir <http://humanum.hypotheses.org/1448> [d.c. 2015/06/12].

Exposer : Partant du constat que de nombreuses équipes de recherche en SHS qui produisent des données scientifiques ne disposent pas de l'infrastructure numérique nécessaire qui permettrait un accès persistant et interopérable à leurs données, la TGIR Huma-Num a conçu et lancé en 2014 un service d'exposition de données (texte, audio, vidéo, etc.). Le service NAKALA s'adresse aux équipes de recherche ayant peu de moyens informatiques pour diffuser elles-mêmes leurs données dans les règles de l'art (en particulier à l'heure du *cloud* et des enjeux d'accès et de confidentialité des données) et dans de bonnes conditions d'interopérabilité.

NAKALA propose deux grands types de services : des services d'accès aux données et des services de présentation des métadonnées. Les producteurs de données numériques ainsi soulagés de la gestion purement technique, peuvent se consacrer à la valorisation scientifique de leurs données. Un identifiant unique est associé à chaque donnée et permet ainsi de citer les données indépendamment de leur localisation réelle : la technologie proposée est basée sur l'attribution d'identifiants de type *handle*, qui possèdent un mécanisme d'adressage générique, sans lien avec l'institution qui porte le service. C'est ce qui rend pérenne l'accès à la donnée, même en cas de changement ou d'évolution de l'institution qui porte le service.

NAKALA dispose d'un entrepôt RDF (*Resource Description Framework*) de type Triple Store, qui permet de partager les informations en utilisant les principes, méthodes et technologies du Web de données. L'utilisation de ces technologies standardisées permet de bâtir des applications de valorisation de ces données, par exemple des cartes interactives. Les données peuvent ainsi être « moissonnées » par des services spécialisés, comme ISIDORE, mais aussi Europeana, Gallica, etc., ce qui permettra d'en accroître la visibilité. La complémentarité avec le service ISIDORE est totale, puisque les deux services partagent les mêmes principes.

Les services numériques de la TGIR sont issus de l'infrastructure « Adonis », financée en 2009 par le plan de relance gouvernemental. La création d'Huma-Num a permis d'en faire un bilan fonctionnel en 2013 entraînant sa restructuration en 2014 afin d'être en capacité d'améliorer les services existants et d'en accueillir de nouveaux. Cette lourde opération a été présentée en conseil scientifique de la TGIR et validée en comité de pilotage en juin 2014. À cette occasion, un bilan financier a été réalisé au regard de l'investissement engagé et des matériels et services mis à disposition par le CC-IN2P3 : il apparaît que les 410 k€ représentent bien précisément la valeur d'usage sur 5 ans.

La nouvelle infrastructure de la TGIR Huma-Num a été déployée en deux phases (2013 : période de transition ; 2014-2015 exploitation de la nouvelle infrastructure) sous la forme de progressions itératives (cluster #1 vers cluster #2 à la fin 2015).

Caractéristique de la nouvelle infrastructure physique et logique (2013-2015).

Cluster #1 :

- Mise en œuvre de 8 serveurs achetés fin 2013, réalisant un hébergement modernisé de l'ensemble des services et première série d'actions pour la bascule de l'infrastructure de 2009 vers la nouvelle dans le courant de 2014.
- Mise à jour des solutions de stockage de données (NAS)
- Objectif de migration complète fin 2015

Cluster #2 :

- Étude comparative et maquettage des solutions de virtualisation de serveurs : écosystème VMware, écosystème KVM (virt-manager, Proxmox, oVirt/RHEL), virtualisation légère (Linux Vserver, OpenVZ, LXC, Docker)
- Choix de l'écosystème KVM plus ouvert et plus utilisé au sein de la communauté Recherche (dont les SHS), et de LXC pour la virtualisation légère
- Présentation de cette étude à un workshop
- Achat et mise en œuvre de 2 serveurs pour démarrer la virtualisation de l'infrastructure
- S1-2015 : passage en production pour une dizaine de machines virtuelles
- S2-2015 :
 - Virtualisation complète des serveurs
 - Capacité d'accueil importante dans les 2 modes (virtualisation lourde et légère)



Mise en production de nouveaux services sur le Cluster #2 :

- Mise en production de services d'éditorialisation/diffusion des données de la recherche dans l'écosystème SHS
- Service de listes de diffusion (logiciel SYMPA) mise en œuvre pour nos utilisateurs
- Documentation détaillée et publicité autour des logiciels commerciaux achetés par Huma-Num pour les communautés SHS : en particulier les licences pour les éditeurs de données de la recherche.

IV - Dimension internationale

La TGIR Huma-Num porte la participation de la France dans l'ERIC DARIAH et participe à 3 programmes H2020 acceptés dans le cadre de son action pour les humanités et le patrimoine numériques de la recherche en SHS et arts. Dans ces 3 programmes H2020, la TGIR Huma-Num fournit soit des services dédiés et issus des savoir-faire maîtrisés par la TGIR, soit de l'expertise sur le domaine des infrastructures de recherche en SHS. Dans ce cadre, la stratégie de la TGIR est de porter ses services numériques dans le cadre européen des infrastructures de recherche dans le respect des modèles de développement des infrastructures nationales des pays de l'Union et dans celui des normes et standards en vigueur dans les données des SHS numériques.

- PARTHENOS (infradev2) : <http://www.parthenos-project.eu> - 388k€ pour la France (partenaires : Huma-Num, INRIA) sur 4 ans.

Le programme PARTHENOS (*Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies*), qui vient d'être accepté dans le cadre d'H2020 et dans lequel Huma-Num est partenaire, vise à renforcer la cohésion de la recherche en SHS dans les disciplines du patrimoine culturel, de l'histoire, de l'archéologie, de la linguistique et plus largement des humanités. Il s'agit aussi de proposer un cluster thématique aux infrastructures européennes de recherche (CLARIN, DARIAH) en vue d'une intégration des outils et méthodes des e-infrastructures et d'autres infrastructures internationales.

PARTHENOS regroupe 15 partenaires européens qui vont travailler pour mettre en commun des standards issus des communautés scientifiques, mais aussi développer des services mutualisés interconnectant soit des services existants chez les partenaires, soit définissant de nouveaux services qui seront co-construits. PARTHENOS s'intéressera plus particulièrement aux différents cycles de vie des données numériques dans les disciplines SHS de son périmètre. Définir, à l'échelle de l'Europe, une politique d'accès aux données, d'interopérabilité, d'archivage et de réutilisation des données est un programme important qui devra se nourrir des expériences/projets des partenaires de PARTHENOS.

PARTHENOS est construit en coordination avec les infrastructures européennes CLARIN et DARIAH, dont il est – pour cette dernière – un projet affilié. Il est coordonné par l'Italie et le Pr. Franco Niccolucci. Le lancement officiel aura lieu au début de l'été 2015 à Florence.

La TGIR Huma-Num développe des services dédiés à PARTHENOS et participe à la rédaction de guide de bonnes pratiques.

- DARIAH-H2020 (infradev3) : <http://www.dariah.eu> - 312k€ pour la France (partenaires : Cléo, Huma-Num, CESR, Atif, MoDyCo, ArScan, CIHAM, Frantiq, IRHT, consortium IRCOM, ABES, ministère de la Culture et de la Communication) sur 2 ans.

Il s'agit d'un appel ciblé, conçu par la CE pour les ERIC nouvellement créés afin d'aider à leur démarrage. Le dossier de réponse a été élaboré à l'automne 2014 par l'ERIC DARIAH auquel étaient associés les trois institutions de pays membres fondateurs de l'ERIC : la France (avec le CNRS & les partenaires nationaux), l'Allemagne (avec l'université de Göttingen) et les Pays-Bas (avec le DANS). Le projet est prévu pour une durée de deux ans et il est organisé autour de huit « *Work-Packages* » (WP). La TGIR Huma-Num a coordonné la réponse française et a sollicité les contributeurs à DARIAH en 2015 en fonction des différents WP affectés à la France par l'ERIC. La France est en charge du WP n°4 (*Training and Education*).

- RITRAIN (infrassupp3) : http://cordis.europa.eu/project/rcn/194941_en.html - 39 000 euros (partenaire : CNRS) sur 4 ans.

Le projet RITRAIN élaborera un programme de formation permettant aux infrastructures de recherche en Europe - dans tous les domaines des sciences - d'acquérir une expertise sur la gouvernance, l'organisation, la gestion financière et le personnel, le financement, la propriété intellectuelle, la prestation de services et de sensibilisation dans un contexte international. Il est conçu et développé par des experts qui ont mis en place et/ou gérés des infrastructures de recherche du concept à maturité. La réponse est pilotée par l'ERIC BBMRI. Le CNRS (via la TGIR Huma-Num) est partenaire dans le cadre d'apports sur les questions d'expertises.

- Coordination de la participation de la France dans DARIAH

La TGIR Huma-Num porte et coordonne la participation de la France dans l'ERIC DARIAH qui est entré – après une longue phase de préparation (2011-2014) dans sa phase de construction (création de l'ERIC le 15 août 2014, inauguration en nov. 2014 à Paris au MENESR). A ce titre, Huma-Num établit la feuille de route nationale, finance la contribution financière annuelle (cash : 151k€) et organise le rassemblement des contributions en nature (in-kind). En 2014, la coordination opérée par la TGIR Huma-Num a été centrée sur le rassemblement des contributions en nature (appel à contributions en 2014 pour 2015 - chiffrage des contributions en nature pour 2015 : 4 millions euros) :

- 51 contributions ont été déposées, toutes les contributions ont été examinées au niveau français par le conseil scientifique de la TGIR, des recommandations aux candidats ont été formulées. Toutes les contributions ont été validées au niveau européen.
- 28 entités contributrices (ABES, Archéovision, Arscan, CCSD, CESR, CHJ, CIHAM, CINES, Cléo, FMSH, FRANTIQ, Huma-Num, INIST, InVisu, IRHT, MAP, MMSH, Modyco, Praxiling, les consortiums COSME, IRCOM, Archives des ethnologues et CAHIER, les Equipex Ortolang et Biblissima, enfin la DREST, la DSIP et la BNF, qui dépendent du ministère de la Culture et de la Communication)
- 27 disciplines SHS représentées
- 243 personnes impliquées ; 232 représentent plus de 33 ETP, issues de 40 établissements dont 23 universités
- Mise en ligne des descriptifs des contributions françaises (version française) sur le site de DARIAH-FR

La participation des communautés françaises reste déterminante dans l'ERIC DARIAH car les contributions françaises, en particulier celles issues des plateformes nationales de premier plan (éditions, archives ouvertes, données, bibliométrie, etc.) sont les principaux fournisseurs de services, de savoir-faire, et d'expertises dans DARIAH.

V - Perspectives

On peut apprécier les perspectives de la nouvelle TGIR au regard de l'opération de fusion du TGE Adonis et de l'IR Corpus, des caractéristiques d'Huma-Num et de sa prospective.

1. Évaluation de la fusion

L'opération de fusion peut être évaluée selon les objectifs initiaux :

- **Objectif 1** : la maintenance et la consolidation des missions et des activités opérées par le TGE Adonis et Corpus IR.

Indicateurs de réussite : reprise complète de l'architecture de l'infrastructure numérique, en cours d'achèvement. Passage d'Isidore au multilinguisme, mise en production de l'archivage à long terme avec le CINES, création de l'ERIC Dariah en 2014, réalisation d'une action nationale de formation par an.

- **Objectif 2** : le développement de nouveaux services dédiés aux données de la recherche.

Indicateurs de réussite : labellisation de deux nouveaux consortiums (Musica et 3D), création du service d'exposition de données Nakala largement utilisé par les communautés SHS, du service d'accueil de projets sur VM, du service de pack Omeka.

- **Objectif 3** : offrir rapidement une bonne visibilité de la nouvelle infrastructure dans les communautés scientifiques SHS.

Indicateurs de réussite : création et diffusion d'une identité propre à la nouvelle structure (nouveau nom, nouvelle identité visuelle, nouveaux sites, présence sur les réseaux sociaux)

2. Caractéristiques actuelles

- Huma-Num appuie son action sur une forte implication des communautés scientifiques SHS pour la définition, l'usage et le suivi de ses services.
- Huma-Num est leader en matière de web sémantique dans le domaine de la recherche en SHS, en particulier en matière de mise en œuvre des méthodes et techniques du web sémantique (web de données) à grande échelle (big data).
- Huma-Num est l'unique partenaire labellisé par les Archives de France pour l'archivage à long terme des données numériques de la recherche (en coopération avec le CINES).
- Huma-Num coordonne la participation de la France dans DARIAH (ERIC).

La TGIR Huma-Num présente une forte valeur ajoutée par rapport aux infrastructures similaires existantes dans son domaine scientifique par les points suivants :

- Définition et mise en œuvre de services numériques tournés vers les données de la recherche en SHS et adaptés aux besoins multiples des communautés SHS (services spécifiques épousant les pratiques de recherche).
- Convention CINES/MCC sur l'archivage à long terme des données numériques
- Mise en œuvre du web sémantique pour les données en SHS (publications, données de la recherche). Dans ce domaine la TGIR Huma-Num est reconnue au niveau mondial pour son expertise en matière de signalement des données de la recherche SHS dans le web sémantique (Linked Open Data) avec les services ISIDORE et NAKALA.
- Articulation de ses services au niveau national, européen et international si applicable

3. Prospective

La TGIR Huma-Num doit poursuivre son ancrage en tant qu'infrastructure de recherche déployant des services numériques forgés au sein des communautés scientifiques des SHS et portées à l'échelle nationale, européenne et internationale. Il s'agit de déployer des services et des outils au meilleur niveau et au bon moment dans les processus de recherche, en tenant compte des modalités d'appropriation et d'utilisation différents selon les disciplines scientifiques.

Il s'agira aussi d'accroître la lisibilité des services, d'en développer/adapter/déployer de nouveaux en co-animation avec les équipes scientifiques et l'équipe de la TGIR. La mise en exploitation sur le long terme sera l'un des enjeux majeurs pour les années à venir. Les programmes H2020 dans lesquels la TGIR est partenaire apporteront des pistes de collaboration internationale qu'il faudra articuler aux services déployés pour les communautés nationales. La TGIR Huma-Num devra anticiper toute croissance à la lumière de ses moyens humains, de la qualité de ses services pour les communautés SHS et de l'écosystème de l'e-science internationale. Sur le plan national, elle devra parachever son réseau distribué en poursuivant et structurant la coopération avec les MSH notamment.

La TGIR Huma-Num se positionne ainsi dans le paysage des grandes infrastructures de recherche internationales. Son assise lui permet de déployer des services dédiés aux données de la recherche à l'international et en coopération avec le réseau d'acteurs SHS et *digital humanities* autour de DARIAH-ERIC et dont il faudra évaluer régulièrement les gains pour les communautés nationales. Les actions de la TGIR Huma-Num s'inscrivent de plus en plus au sein de coopérations internationales non-européennes (via la Research Data Alliance) avec des partenaires universitaires du monde francophone (dont l'Université de Montréal) et non-francophone avec des collaborations déjà initiées avec des réseaux européens tel que le REDIAL/CEISAL, Calcul Québec, Université de Stanford (en particulier autour des projets ISIDORE, NAKALA).

VI - Liste des consortiums au 30 juin 2015

Les consortiums réunissent plusieurs unités et équipes autour de thématiques et d'objets communs pour lesquels ils définissent des procédures et standards numériques partagés. Ils formulent un programme de numérisation ou de documentation pluriannuel pour lequel ils s'engagent à apporter une partie des moyens nécessaires (en particulier en personnel), et pour lequel ils demandent, en outre, un soutien de la part de la TGIR. Afin d'obtenir ce soutien, les consortiums doivent être labellisés.

Consortium Musica

Labellisé en 2014

Le consortium Musica réunit plusieurs unités et équipes de recherches autour d'une thématique commune : la valorisation et l'exploitation de données musicales numériques par l'utilisation de référentiels communs, la diffusion de bonnes pratiques d'encodage et des opérations de numérisation des données.

L'un des objectifs de Musica sera de coordonner la réflexion épistémologique des chercheurs, ainsi que le développement de nouveaux outils de représentation et d'analyse de la musique. Ces outils pourront prendre appui sur les langages HTML5, CSS3 et javascript, qui constituent les nouveaux standards de l'internet et permettent de réaliser des dispositifs multimédia interactifs facilement adaptables d'un projet à l'autre. Une attention particulière sera portée aux manières d'intégrer ces nouveaux outils de représentation à ceux développés dans les axes du consortium, notamment la constitution de corpus de partitions et le développement de plateformes d'archives sonores en ligne.

En savoir plus : <http://humanum.hypotheses.org/503>

Consortium 3D

Labellisé en 2014

Les usages des technologies 3D se diversifient et occupent une place de plus en plus importante dans les processus d'étude des sociétés humaines. Plusieurs équipes de recherches jouent un rôle dans l'appropriation de ces nouveaux outils et collaborent ainsi à l'émergence de nouvelles pratiques. Le consortium 3D se propose de fédérer ces pratiques en les adossant à un réseau d'unités fortement impliquées, afin d'optimiser l'insertion des méthodologies 3D au service d'objectifs de recherche, tant au niveau régional, national, qu'international.

En savoir plus : <http://humanum.hypotheses.org/480>

Consortium Mémoires des archéologues et des sites archéologiques [MASA]

Labellisé en 2012

Ce consortium réunit un certain nombre d'acteurs nationaux déjà rassemblés dans le réseau national des Maisons des Sciences de l'Homme (GIS RMSH) traitant d'archéologie et le Musée d'archéologie nationale.

En vue de favoriser toute recherche nationale ou internationale, ce consortium propose la mise en place d'un accès unifié à des corpus numérisés avec des procédures et des outils documentaires et archivistiques communs.

En savoir plus : <http://masa.hypotheses.org>

Consortium Sources Médiévales (COSME)

Labellisé en 2012

Le positionnement de ce consortium est clairement pluridisciplinaire : il s'appuie sur une communauté scientifique largement ouverte à toutes les disciplines, rassemblant archéologues, historiens, historiens de l'art, philosophes, linguistes, littéraires... et réunie autour d'une pratique commune : celle de la recherche sur le passé médiéval par le biais des sources les plus diverses. Ces sources sont au cœur des préoccupations du consortium : projets d'édition ou de transcription, de reproduction numérique ou d'un répertoire les recensant.

En savoir plus : <http://cosme.hypotheses.org>

Consortium Cartes et photographies pour les géographes

Labellisé en 2012

En France, quelques portails cartographiques et plateformes de diffusion de données et métadonnées ont été mis en place, principalement par des institutions, mais il est nécessaire de développer ce travail à une autre échelle afin de généraliser l'accès à d'autres fonds pertinents et d'améliorer la diffusion des images géographiques. L'objectif de ce consortium est donc de rendre accessibles, consultables et mobilisables des données cartographiques et photographiques nombreuses et éparses, qui constituent des fonds de laboratoires de recherche, de bibliothèques remarquables ou des fonds de chercheurs, ressources en général sous-exploitées faute d'accès en ligne ou de documentation suffisante.

En savoir plus : <http://humanum.hypotheses.org/191>

Consortium Archives des sciences sociales du politique (Archipolis)

Labellisé en 2012

Ce consortium a pour mission d'archiver les recherches engagées à l'interface des sciences sociales et portant sur les différentes composantes de l'objet politique – des plus spécifiques telles que les élections, les partis, les syndicats et autres organisations politiques, le métier politique, l'action publique, les institutions, la citoyenneté, mais aussi plus largement toutes formes collectives d'appartenances et de comportements mettant en jeu les rapports de pouvoir.

En savoir plus : <http://archipolis.hypotheses.org> et <http://humanum.hypotheses.org/147>

Consortium Archives des mondes contemporains

Labellisé en 2012

Ce consortium a pour objet la création d'un instrument de recherche collectif qui permette de répondre aux problèmes qui découlent de la nature même des archives de la période contemporaine (XIXe-XXe siècle) qui sont caractérisées tout à la fois par leur masse, l'hétérogénéité des supports (du papier au microsillon en passant par la photographie ou les disquettes informatiques, CD-rom et les supports vidéo anciens et récents), leur fragilité, la diversité de leur statut juridique (public ou privé), et la dispersion des lieux où elles sont conservées.

En savoir plus : <http://arcmc.hypotheses.org> et <http://humanum.hypotheses.org/231>

Consortium CAHIER - Corpus d'Auteurs pour les Humanités : Informatisation, Édition, Recherche

Labellisé en Septembre 2011

CAHIER est un consortium transversal interdisciplinaire. Il a pour objectif de fédérer les différentes initiatives existantes ou en projet en France dans les domaines des "corpus d'auteurs, qu'ils relèvent de la littérature, de la philosophie ou d'une thématique liée à une école ou à une pratique, afin d'apporter coordination, partage d'expériences et de favoriser l'accès aux données.

Les corpus pris en compte dans ce consortium sont fortement associés à une activité éditoriale, qu'elle soit seulement numérique ou qu'elle ait un double support (sur papier et en ligne). CAHIER se définit par rapport à l'existence d'une œuvre (incluant les documents préparatoires), ou de plusieurs œuvres identifiées, dont la cohérence mérite d'être soulignée, publiée et outillée pour donner lieu à de nouvelles recherches.

En savoir plus : <http://cahier.hypotheses.org/>

Consortium Corpus Oraux et Multimodaux (IRCOM)

Labellisé en 2011

Ce consortium a vocation à fédérer les équipes, laboratoires, chercheurs et enseignants-chercheurs engagés dans la constitution de corpus oraux et multimodaux, afin de faire converger les pratiques et de les rendre conformes aux standards internationaux.

En savoir plus : <http://ircom.huma-num.fr> et <http://humanum.hypotheses.org/184>

Consortium Corpus Écrits

Labellisé en 2011

Ce consortium se donne précisément pour but de fédérer les équipes et laboratoires, les chercheurs, enseignants-chercheurs, ou ingénieurs engagés dans la production de corpus numériques écrits, quels que soient la langue et l'alphabet considérés et d'offrir la représentation la plus large possible de cette communauté, afin d'accompagner le développement des corpus écrits, d'en faire converger les pratiques et les besoins, et de financer des actions répondant à ses missions. Il vise ainsi à répondre aux besoins d'information, de formation, de partage des bonnes pratiques et de diffusion des standards européens et internationaux en matière de création, production, développement et valorisation des corpus numériques écrits.

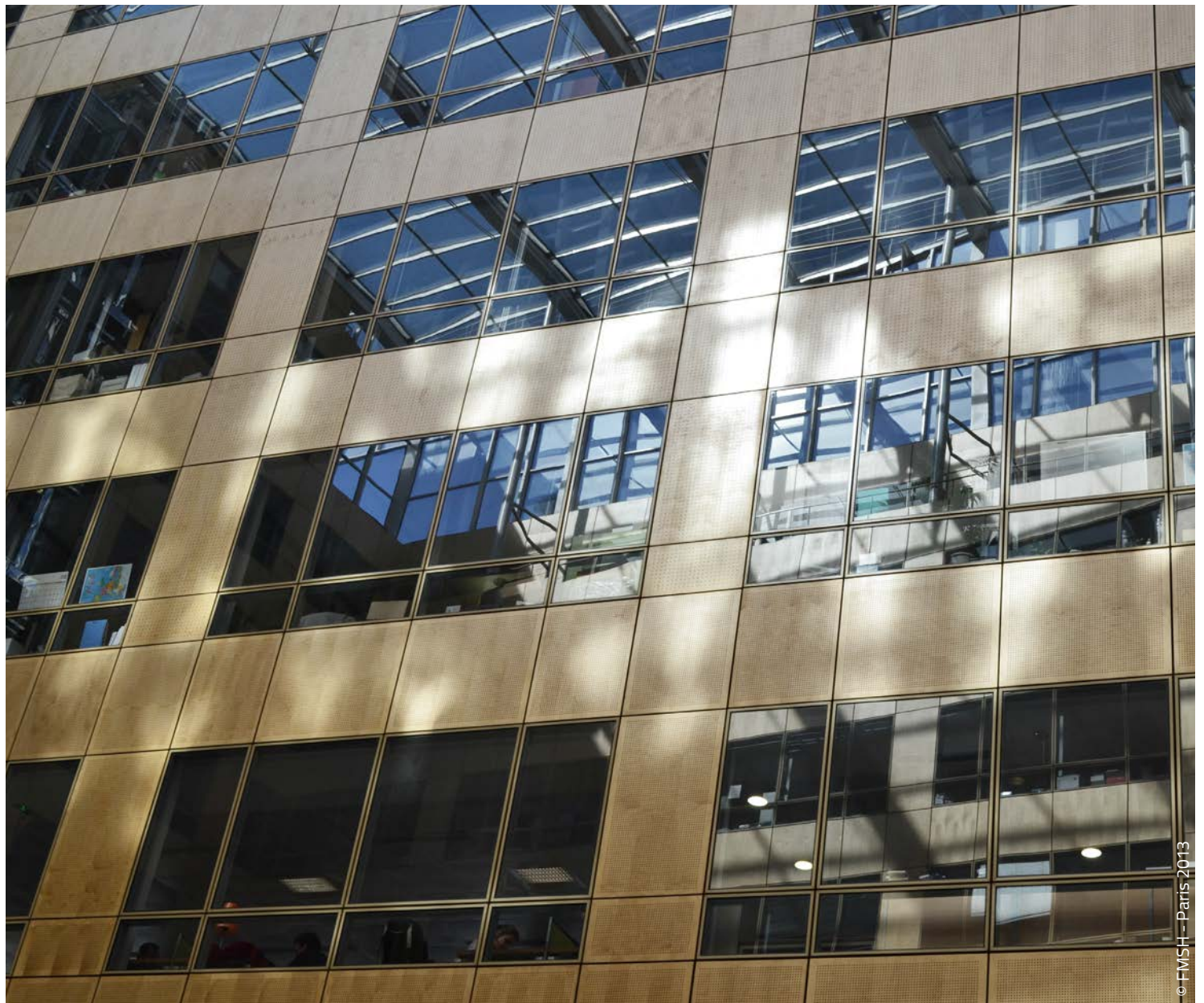
En savoir plus : <http://corpusecrits.huma-num.fr> et <http://humanum.hypotheses.org/315>

Consortium Archives des ethnologues

Labellisé en Juillet 2011

Ce consortium est créé pour répondre aux questions spécifiques qui se posent aux ethnologues devant la singularité de leurs corpus, l'importance théorique et patrimoniale de leurs travaux, la richesse et la diversité humaine qu'ils représentent. Il s'agit, dans un premier temps, d'assurer la conservation et la pérennisation des données collectées par les ethnologues sur le terrain ainsi que de tous documents pouvant aider à la contextualisation de ces données ou à une meilleure compréhension de l'élaboration des travaux anthropologiques. Dans un deuxième temps, l'objectif du consortium est de mettre à disposition ces matériaux dans le respect de la propriété intellectuelle et des normes éthiques.

En savoir plus : <http://ethnologues.huma-num.fr>



© FMSH - Paris 2013

Retrouvez toute l'actualité d'Huma-Num sur :
<http://www.huma-num.fr>
<http://humanum.hypotheses.org>

