



**Programme thématique en sciences
humaines et sociales**

**CORPUS ET OUTILS DE LA RECHERCHE
EN SCIENCES HUMAINES ET SOCIALES**

APPEL A PROJETS DE RECHERCHE 2006

Destiné aux différentes disciplines des
Sciences humaines et sociales

Date de publication : 14 mars 2006

Date limite d'envoi des dossiers : **12 mai 2006**

Les dossiers doivent être envoyés par courrier électronique et par courrier postal
en 4 exemplaires (un original et 3 copies) cachet de la poste faisant foi

à
Programme corpus et outils de la recherche
en sciences humaines et sociales
ENS LSH
15 Parvis René Descartes
BP 7000
69342 Lyon

RENSEIGNEMENTS SCIENTIFIQUES
pierre-olivier.pin@agencerecherche.fr

Formulaire de soumission et renseignements sur <http://www.agence-nationale-recherche.fr/>

■	CHAMP DE L'APPEL A PROJETS	4
■	ASPECTS EPISTEMOLOGIQUES ET METHODOLOGIQUES	4
■	CORPUS ET ARCHIVES EN SCIENCES HUMAINES ET SOCIALES	5
■	CORPUS ET ANALYSE DES LANGUES ET DES CULTURES	5
■	GESTION ET ANALYSE DES LA DOCUMENTATION ELECTRONIQUE EN SCIENCES HUMAINES ET SOCIALES	6

CORPUS ET OUTILS DE LA RECHERCHE EN SCIENCES HUMAINES ET SOCIALES

Le recours à ce que l'on nomme commodément « corpus » s'est généralisé dans de nombreux secteurs des sciences humaines et sociales : sciences du langage, de l'information et de la communication, littérature, histoire, sociologie, psychologie, économie, archéologie...

En réalité, le terme « corpus » peut désigner deux stades documentaires différents : d'une part, un ensemble de sources, écrites, orales ou audiovisuelles brutes, de l'autre, l'information élaborée à partir de ces sources. Dans certains cas, comme en linguistique, la méthodologie employée permet d'analyser le document brut, mais ce mode d'emploi est peu fréquent ; le plus souvent, une étape de la recherche doit être consacrée à la transformation des archives en corpus. L'objectif de cet appel d'offre est de répondre à une demande des chercheurs désireux d'obtenir une aide pour un projet de recherche nécessitant une collecte documentaire d'un projet de recherche dûment argumenté.

Les candidats qui répondront à cet appel d'offre seront pris en considération dans la mesure où ils respecteront les exigences suivantes :

1. La réalisation des projets de recherche devra nécessiter la construction de documentation et d'information.
2. Le projet scientifique devra être justifié dans le cadre de la discipline (ou de plusieurs disciplines), tant sur le plan méthodologique et épistémologique qu'en termes d'originalité.
3. Les outils de traitement du corpus envisagé devront être détaillés et les conditions de leur élaboration précisées.
4. Les candidats devront détailler le contenu des informations et justifier les choix opérés aussi bien sur le plan qualitatif que quantitatif, en précisant les méthodes mises en œuvre pour critiquer la source et assurer la représentativité des données.
5. En particulier, lorsque la collecte de données représentera un investissement lourd, les candidats devront le justifier par les possibilités de partage des outils et des données au sein de la communauté scientifique.
6. La production et l'organisation des données devront respecter les normes agréées par les communautés scientifiques pour assurer leur accès à la fois ouvert et sécurisé, en particulier les normes légales en matière de protection des données individuelles.
7. Un calendrier précisera les étapes du projet.

CHAMP DE L'APPEL

L'appel à projets concerne le financement de projets de recherche qui nécessitent la mise en place, la valorisation, l'enrichissement, la documentation de corpus, ainsi que l'amélioration des outils et des procédures d'analyse.

L'appel vaut pour tous les types de sources (inscriptions, textes, images, documents sonores, documents multimédia, données statistiques...) et concerne l'ensemble des sciences humaines et sociales.

L'appel ne couvre pas les activités de stockage et de numérisation sans objectif de recherche précis. Il n'a pas pour objet la simple constitution de fonds documentaires ou d'édition numérique.

L'évaluation des projets soumis sera particulièrement attentive aux retombées du projet pour d'autres champs et d'autres chercheurs que les équipes financées. Au terme d'une période d'exclusivité raisonnable, les données devront être accessibles à l'ensemble des scientifiques intéressés, suivant des modalités qui pourront varier selon les disciplines ou selon les paradigmes.

Les projets pourront mobiliser des approches interdisciplinaire et comparatives.

Parmi les thèmes qui pourront être abordés, nous proposons un certain nombre de directions de recherche.

ASPECTS EPISTEMOLOGIQUES ET METHODOLOGIQUES

Cette première direction de recherche aborde la notion de corpus sous un double aspect épistémologique et méthodologique. Elle concerne en particulier les problématiques suivantes :

- stratégies d'interrogation, d'extraction, de classification des données d'un corpus et questions relatives aux principes et pratiques de catégorisation ;
- recherches en typologie des textes et des discours, des sources iconographiques ou sonores nécessaires au travail de catégorisation des corpus ;
- contraintes d'usage et d'exploitation des bases et des corpus (droits juridiques afférents aux documents initiaux et enrichis, protection des personnes et des libertés publiques, protection de la propriété intellectuelle...);
- normalisation assurant un partage effectif des données entre chercheurs de disciplines ou de nations différentes et garantissant la reproductibilité des collectes et des analyses ;
- méthodologies de traitement des données biographiques et spatiales.

CORPUS ET ARCHIVES EN SCIENCES HUMAINES ET SOCIALES

La contribution des archives à la recherche permet d'éclairer le statut de la preuve en sciences humaines et sociales et d'approfondir le processus de création scientifique en s'intéressant à des méthodologies et des cadres interprétatifs précis. Peuvent être évoquées notamment :

- la description analytique des inventaires archivistiques, leur accès à distance, leur analyse assistée ;
- la muséographie (les musées comme corpus et archives constituant une ressource permanente de la recherche) ;
- la reconstitution par synthèse d'images ou par synthèse vocale d'univers spatiaux et langagiers, notamment sur la base de données géo-archéologiques, paléoenvironnementales, géohistoriques, linguistiques ;
- le traitement des données paléographiques, codicologiques, graphématiques, audiovisuelles ;
- l'archivage de données électroniques dématérialisées.

CORPUS ET ANALYSE DES LANGUES ET DES CULTURES

Cette troisième direction de recherche propose de développer des projets consacrés à la description et/ou à la modélisation des langues et des cultures (occidentales et non occidentales) à partir de matériaux et de documents textuels, sonores, iconographiques, vidéo, multimédia.

Elle concerne les problématiques de recherche suivantes :

- relations entre les théories linguistiques et les modes de production des corpus ;
- spécificité des corpus de langue parlée qui se plient difficilement à la normalisation de l'écrit (langues non inventoriées ou en voie d'extinction, pratiques orales des jeunes enfants, discours multilingues, contextualisation et multimodalité) ;
- articulation entre les approches qualitatives et quantitatives des corpus oraux et écrits ;
- constitution de corpus permettant d'étudier l'évolution des langues et leurs caractéristiques typologiques ;
- linguistique de corpus et didactique des langues ;
- problèmes théoriques de la traduction ;
- édition scientifique de corpus textuels permettant une avancée de la recherche dans la science des textes et dans l'histoire des idées scientifiques ;
- exploitation statistique, lexicométrique ou terminologique des corpus textuels ;
- étude des cultures fondée sur l'exploitation de corpus multidisciplinaires ;
- recherches sur la genèse des textes.

DONNEES NUMERIQUES EN SCIENCES HUMAINES ET SOCIALES

Au nombre des nouvelles ressources utiles à la compréhension de la société figurent les progrès dans la mise à disposition des données statistiques publiques, le développement des grandes enquêtes, la mise en ligne de catalogues de données à l'échelle internationale, la constitution de bases de données longitudinales ou la généralisation de la numérisation des données de gestion. Mais on ne peut créer, enrichir ou exploiter des corpus numérisés qu'à condition de les analyser selon des méthodes rigoureuses et de mobiliser d'importants moyens techniques et financiers.

Le présent appel à projets encourage les démarches scientifiques qui s'appuient sur des corpus de données numériques et de textes ou images numérisés, en privilégiant :

- l'apparition de travaux originaux mobilisant des sources inédites ou mal connues, combinant des informations jusque là dispersées (issues par exemple de secteurs disciplinaires différents) ou actualisant des sources anciennes, testant des méthodologies innovantes ou enrichissant notablement des sources existantes ;
- les processus de mutualisation permettant à des équipes de valoriser des corpus par une utilisation croisée ou récurrente, comme c'est le cas des bases de données longitudinales.

Peuvent être concernés les données strictement numériques, les textes, images et sons numérisés, ou les données mixtes nécessaires au développement des recherches en sciences humaines et sociales (provenant, par exemple, de l'enregistrement vidéographique de situations d'interaction ou d'interlocution).