

CNRS — DIS-H&S — Centres de ressources numériques

Lettre de mission

Suite à la création de la direction de l'information scientifique (DIS) et considérant qu'un de ses rôles essentiels est la mise en œuvre de centres de ressources numériques (CRN), il est décidé de créer un tel centre sur le thème des corpus oraux (corpus phonologiques, corpus audio pour la reconnaissance de la parole et corpus d'oral spontané). Le nom de ce CRN est « *Centre de ressources pour la description de l'oral* » et son sigle « *CRDO* ».

Le but général de la mise en place des centres de ressources numériques est de *dynamiser* la recherche Grâce à une l'infrastructure permettant la production, la validation et la diffusion de ressources numériques (la notion de ressource couvre ici aussi bien les *données* que des *outils* permettant de les traiter) et en facilitant l'accès aux services disponibles grâce à une centralisation des connaissances et des moyens. À terme, cette structuration devra faciliter une meilleure transmission de l'information scientifique qui ne soit plus seulement fondée sur les publications, et permettre une veille portant sur l'observation des ressources considérées ainsi que des actions de formation visant à faire connaître et admettre les règles émergentes, pour que tous ces travaux soient cohérents et compatibles entre eux.

1. Missions du CRN *CRDO*

Les missions générales définies pour ce centre sont les suivantes (précisées dans l'annexe jointe) :

1. « Entrées » : contrôle et acceptation des ressources ;
2. « Stockage » : stockage, maintenance et sauvegarde des ressources ;
3. « Gestion des ressources » : partage, conservation et enrichissement de ressources ;
4. « Administration » : administrations des ressources et aides aux utilisateurs ;
5. « Pérennisation » : documentation et pérennisation par mise à jour et évolution des logiciels et des supports informatiques ;
6. « Accès » : récupération des ressources, aide et réponse aux utilisateurs, gestion des protections juridiques des données et ressources en privilégiant autant que faire se peut des licences de type « ressources libres » ;
7. « Formation » : formation des producteurs et des utilisateurs à l'utilisation du centre et aux méthodologies d'annotation, de codage et de normalisation.

2. Mise en place du CRN *CRDO*

Ce centre résultant de la mise en commun des propositions de l'UMR 7107 (LACITO – Langues et Civilisations à Tradition Orale) et de l'UMR 6057 (LPL - Laboratoire Parole et Langage), dans la phase de démarrage, le centre sera adossé à ces deux unités, l'objectif étant à terme de l'organiser sous forme d'une structure distincte.

1. Mise en place d'un comité scientifique et d'un comité technique du *CRDO*. Le rôle du comité scientifique sera de donner son avis sur les diverses tâches réalisées par le centre de ressources numériques ; le rôle du comité technique sera d'assurer le suivi au

jour le jour de ces opérations. *Noms à proposer, avalisés par la DIS et le département H&S.*

2. Création d'un site web spécifique avec, si possible, un domaine spécifique comme : www.crdo.fr et des adresses spécifiques : contact@crdo.fr (démarches à effectuer auprès de Renater pour réserver le domaine correspondant).
3. Affectation explicite au futur *CRDO* de personnels chargés de la gestion de ressources, actuellement affectés aux unités LACITO et LPL].

3. Premières actions

1. Réflexion sur le ou les domaines les plus pertinents (la langue française et les langues de France ou les langues sur lesquelles travaillent les chercheurs français)
2. Recensement et catalogage de l'existant dans le domaine des corpus oraux, en termes de données, de logiciels, de laboratoire partenaires et d'opérations internationales analogues.
3. Analyse approfondie des articulations possibles avec le MPG (Allemagne) qui a déjà une structure d'archivage importante et efficace dans ce domaine.
4. Activation d'un site web dynamique / portail pour le centre à l'adresse www.crdo.fr. La première version du site devra être opérationnelle au terme des trois premiers mois suivant l'officialisation du centre.
5. Visibilité sur le site du centre des métadonnées correspondant à l'ensemble des ressources des laboratoires partenaires. Ces métadonnées devront suivre autant que possible les normes et recommandations internationales préconisées.
6. Mise en ligne des ressources actuellement disponibles et libres de droit et valorisation des corpus mis en ligne par publication [semi-]automatique sur HAL.
7. Mise en place d'une école de printemps ou d'automne sur l'utilisation du site et sur les aspects codage et normalisation de ressources.

4. Actions futures

1. Définir un appel à projets dans le domaine des corpus oraux pour l'extension des données existantes, la numérisation de nouvelles données et le développement de processus de traitement. La sélection des projets sera réalisée par le comité scientifique, la DIS et le département H&S.
2. Viser la réalisation des sept fonctions définies plus haut.
3. En particulier, approfondir la question du dépôt des originaux, du stockage des documents numériques et de leur migration, de la pérennité des supports. Cet aspect doit être discuté avec les autres centres de ressources numériques et géré en collaboration avec les autres institutions à vocation d'archivage et de conservation.