

Introduction à CLARIN

www.clarin.eu

Université Paris Diderot

5 septembre 2017

Bente Maegaard

Université de Copenhague & Vice Executive Director CLARIN
ERIC

bmaegaard@hum.ku.dk



CLARIN en quatre points

- CLARIN est l'INfrastructure Commune de Ressources Linguistiques et technologiques. CLARIN est un ERIC (un consortium de pays)
- qui offre un accès facile et durable aux chercheurs en **les sciences humaines et en les sciences sociales et au-delà**
- aux **données de langage numériques** (sous forme écrite, parlée, vidéo ou multimodale),
- et des **outils avancés** pour les découvrir, explorer, exploiter, annoter, analyser ou combiner, n'importe où qu'elles se trouvent.

CLARIN vision et mission

Vision: Toutes les ressources et tous les outils numériques pour les langues de toute l'Europe et au-delà bénéficient d'un accès unique dans l'environnement en ligne afin de faciliter le travail des chercheurs en sciences humaines et sociales.

Mission: Créer et maintenir une infrastructure qui permet le partage, l'utilisation et la durabilité des données de langage et des outils de recherche en sciences humaines et sociales.

Proposition de valeur: CLARIN rend disponibles des ressources de langage numériques aux universitaires, chercheurs, étudiants et citoyens-chercheurs de toutes disciplines, en particulier dans les sciences humaines et sociales, via l'accès unique. CLARIN offre des solutions à long terme et des services technologiques pour le déploiement, la connexion, l'analyse et le maintien de données et d'outils de langue numériques. CLARIN soutient les chercheurs qui veulent s'engager dans une recherche de pointe et basée sur des données, contribuant à une zone de recherche européenne qui s'avère multilingue.

Commentaires aux déclarations de vision

- *Disciplines:* Même si l'origine de CLARIN est dans les sciences humaines et les sciences sociales, il convient de noter que les offres de CLARIN peuvent être pertinentes pour de **nombreuses autres disciplines où la langue joue un rôle.**
- *Science ouverte:* Plus généralement, CLARIN ne se considère pas comme une installation autonome, mais plutôt comme un acteur dans la réalisation de la vision de **European Open Science Policies**, les nouvelles politiques émergentes, qui relient les chercheurs à travers les frontières nationales et disciplinaires en offrant un accès transparent à des données et services.
- *Les parties prenantes:* La construction et l'exploitation d'une infrastructure impliquent de **nombreux intervenants différents**, chacun avec leurs propres intérêts et attentes: des chercheurs individuels, des instituts de recherche, des archives de données, des fournisseurs de services d'infrastructure, des organismes de financement et des gouvernements, pour n'en mentionner que quelques-uns.

CLARIN en rôles et instruments

CLARIN soutient

- Ceux qui utilisent (ou qui souhaitent ou devraient être encouragés à utiliser) l'infrastructure (**les chercheurs**)
- Ceux qui le peuplent avec des données et des services (**les fournisseurs** - peuvent aussi être des **chercheurs**)
- Ceux qui construisent et exploitent l'infrastructure technique (**opérateurs**)
- Ceux qui gèrent l'infrastructure (**gestionnaires**)

Par une infrastructure qui relie des

- Centres techniques
- Dépôts de données
- Instruments de partage des connaissances

CLARIN en types de données

- Textes littéraires
- Documents parlementaires
- Archives de journaux
- Collections d'histoire orale
- Données orales et vidéo
- Dictionnaires
- Lettres historiques
- Bibliothèques disciplinaires
- Données d'archives institutionnelles
- Archives de diffusion
-
- Accès ouvert, dans la mesure possible
- Accès gratuit pour les universitaires à certains documents protégés par des droits d'auteur
- Toutefois: les lois sur les droits d'auteur limitent l'utilisation de la recherche

Virtual Language Observatory de CLARIN

- Language Resource facet browse: vlo.clarin.eu
- 900,000 records (dont 22,000 francais)
- Qui peuvent être filtrés par: [Language](#) | [Collection](#) | [Resource type](#) | [Modality](#) | [Genre](#) | [Other](#) »



ABaC:us –
Austrian Baroque Corpus

CLARIN pour les chercheurs

- Trouver et déposer des ressources
 - Chercher par métadonnées, chercher par contenu
 - Conservation à long terme, PID (Persistent Identifier)
- Outils avancés et installations informatiques
- Connexion fédérée - accès facile
- Expertise pour les chercheurs
 - Helpdesks, Centres de connaissances, Comités
- Collaboration internationale
- Tutoriels en ligne et exemples d'utilisation
 - Toujours en cours
- Ateliers et séminaires, bourses de mobilité
- Prise en charge des plans de gestion des données

CLARIN pour les institutions et les instituts

- Accès garanti aux services centraux CLARIN
 - Fédération des fournisseurs de services (connexion unique)
 - Services PID
 - Accès aux services EUDAT
 - Services pour métadonnées, registre des centres, statistiques d'utilisateurs (piwik)
- Participation aux projets de l'UE
 - CLARIN ERIC dans des projets (CLARIN-PLUS, PARTHENOS, EUDAT2020, LT Observatory (fini), EUROPEANA-DSI). Plus viendra (par exemple EOSC Hub)
- Accès à l'expertise pour les institutions et les instituts

CLARIN pour membres CLARIN ERIC (pays, ministères)

- Meilleure position de la langue nationale, visibilité
- Visibilité du contenu culturel
- Le coût peut être partagé à travers des frontières, le logiciel est partagé, etc.
- Influence sur la stratégie
- Accès à l'expertise pour les nouveaux pays (ateliers, documents de bonnes pratiques, bourses de mobilité)

Information sur CLARIN pour les nouveaux membres

- Site CLARIN www.clarin.eu
- <https://www.clarin.eu/content/information-potential-new-members>
 - Statuts
 - Proposition de valeur
 - Documents de meilleures pratiques
 - Aperçu des ateliers organisés
 - Aperçu des cotisations annuelles
 - Formulaire de candidature
- Ceci est encore en cours et nous apprécions les commentaires!
- Nous avons des bourses de mobilité spécifiques pour les membres nouveaux et à venir jusqu'à la fin de l'année 2017!

Comment créer un consortium national

- La France est déjà un observateur
- Statuts <http://hdl.handle.net/1839/00-DOCS.CLARIN.EU-75>
 - Nous sommes en train de changer les statuts, mais le changement ne sera pas très important
- Un consortium national peut comprendre une ou plusieurs institutions – vous avez déjà HUMA-NUM
- Il faut avoir un coordinateur national pour CLARIN

Consortium continué, discussion générale

- Quel est un bon consortium?
 - Les institutions importantes de votre pays
 - Couvrir les différents domaines
 - Recherche
 - Ressources linguistiques (universités, bibliothèques, patrimoine culturel, ...)
 - Savoir-faire technique
- Au moins un centre technique CLARIN
- Un très grand consortium peut être trop difficile à gérer. Vous pouvez collaborer avec des institutions en dehors du consortium.

Contribution nationale

- Statuts Art. 3 Contribution aux objectifs et obligations globaux et spécifiques
 - Cotisation annuelle
 - Agir conformément aux statuts (articles 2.2 et 6.2)
 - Consortium national et coordinateur national
 - Contributions nationales spécifiques
 - Coordonnées par CLARIN ERIC
 - Coordonnées localement, conforme à CLARIN ERIC

Contributions nationales coordonnées par CLARIN ERIC

- Contributions coordonnées par CLARIN ERIC
 - a) Centres techniques, au moins un centre de type A ou B
 - b) Accès: ouvert à tous les chercheurs et étudiants dans les pays membres par l'authentification
 - c) Système d'authentification et d'autorisation d'utilisateur
 - d) Infrastructure de partage des connaissances
 - e) Participation dans les comités CLARIN ERIC et toute autre contribution se prêtant à la coordination

Contributions nationales coordonnées au niveau national

- Contributions coordonnées au niveau national
 - a) Augmentations et améliorations des ressources, des outils et des services existants
 - b) Création de nouvelles ressources, outils et services
 - c) Mesures visant à promouvoir l'adoption et l'adhésion des normes CLARIN dans les programmes et projets nationaux
 - d) Mesures visant à créer une infrastructure de partage des connaissances au niveau national
 - e) Toute autre contribution concernant les objectifs de CLARIN

Un peu sur la situation danoise

- DK est un membre fondateur de CLARIN – depuis 2012
- Nous avons un centre technique, à l'Université de Copenhague
- Le consortium danois comprend 4 universités et notre bibliothèque nationale. Il y a encore 2-3 universités qui pourraient rejoindre le consortium
- Le consortium danois est DIGHUMLAB, un réseau d'infrastructures financé par le ministère – un peu comme HUMA-NUM
- DK a participé dans plusieurs projets UE: CLARIN-PLUS, PARTHENOS, LT Observatory
- DK a beaucoup bénéficié de la collaboration européenne

Quels activités pour les chercheurs à CLARIN-DK?

- Enquête
- Séminaires de sensibilisation
- Ateliers
- Participation à des activités de recherche dans le domaine des sciences humaines
- Projets pilotes / prototypes

CLARIN-DK enquête 2012

- Difficile d'obtenir des informations utilisables sur la base de questions dans un questionnaire
- Les chercheurs sont très différents
 - Les nouveaux chercheurs des sciences humaines numériques (DH) n'ont pas de «langage de concepts» pour parler de DH et ont besoin d'exemples, ce qui leur permet d'imaginer de nouvelles façons de faire leur recherche.
 - Les chercheurs qui viennent de commencer dans DH, ont besoin de conseils axés sur les progrès qu'ils ont déjà réalisés, donc ils ont besoin de dialogue et de soutien
 - Les chercheurs expérimentés utilisant déjà des outils et des applications dans leur recherche ont des expériences et souvent des suggestions concrètes.

Ateliers – CLARIN-DK

- CLARIN-DK a organisé plusieurs ateliers / groupes de discussion, par ex.
 - Sur TEI
 - Sur l'annotation (l'annotation manuelle, des outils pour annotation)
 - Etc.
- C'est un bon moyen de commencer les discussions avec les chercheurs, mais il est difficile de faire venir les bons chercheurs - comment savent-ils qu'il y a quelque chose pour eux?

Projets pilotes/prototypes

Dans CLARIN-DK, nous avons trouvé qu'il était très inspirant de développer avec les chercheurs des prototypes qui peuvent être développés encore ou qui fonctionnent comme des cas de test.

- Chansons populaires danois dans CQP: <http://cst.dk/dighumlab/DFKalle/frames-cqp.html> (cliquez par exemple sur "Fréquences")
- Aide à des textes latins: [Locus Classicus](#) - outil d'enseignement du latin
- Københavns Diplomentarium-demo ([description](#)) - une carte est ajouté montrant des rues mentionnée dans les textes

Autres activités au Danemark

A part les activités visées aux chercheurs

- Créer le centre B
- Déposer les données
- Créer les outils
- Participer aux comités
- Créer un centre de connaissance pour le danois
- Etc.

Pour conclure

- CLARIN a une très bonne couverture européenne et continue de croître
- Accès aux données et aux outils, accès aux services de dépôt
- Accès à l'expertise
- Les données CLARIN sont des matériaux de langage – mais pas pour les linguistes, surtout pour la littérature, l'histoire, sociologie etc. – les sciences humaines numériques (DH)
- CLARIN prend en charge le multilinguisme
- CLARIN prend en charge l'accès ouvert et 'open source'
 - CLARIN soutiens des efforts vers des règles de droit d'auteur plus détendues pour la recherche

Au revoir et à bientôt @

www.clarin.eu

ou

bmaegaard@hum.ku.dk

