

CORPUS WEB

ENTRE OUTIL DE RECHERCHE
ET OBJET DOCUMENTAIRE

Cynthia Pedroja – DRIS, Sciences Po)
47^e congrès ADBU – 18/10/2017



CONTEXTE

Développement des services à la recherche

- *Direction des ressources documentaires* (DRD) devient en **2015** *Direction des ressources et de l'information scientifique* (**DRIS**)
- archive ouverte SPIRE, gestion des données de la recherche, iconographie, numérisation, **corpus web**

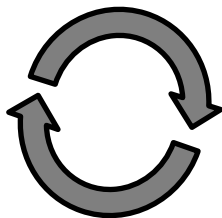
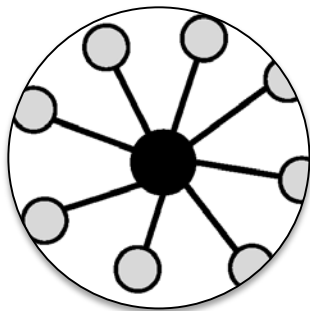
Le web bouleverse la constitution des fonds des bibliothèques

- Comment rendre compte de l'actualité en ligne ?
- Comment dépasser la liste des ressources web ?

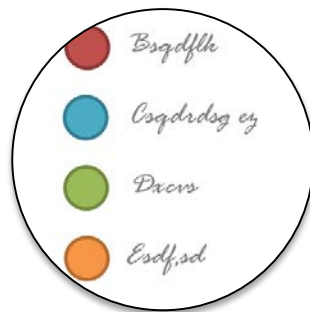
DRIS souhaite répondre à la double problématique à travers les corpus web dans le cadre d'une collaboration avec le **médialab** et des centres de recherche.

Etapes :

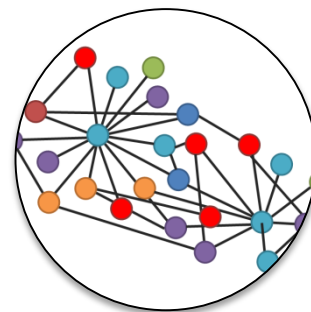
CONSTRUIRE



EXPLORER



EXPOSER



Activités :

- Définir le cœur de corpus : *lister les sites de départ*
- Crawler : *sélectionner les sites pertinents*

- Créer l'outil intellectuel : *déterminer catégories et référentiels*
- Enrichir les sites retenus : *catégoriser*

- Créer le graphe (réseau) : *cartographier*
- Outiller le corpus : *manipuler par filtre*

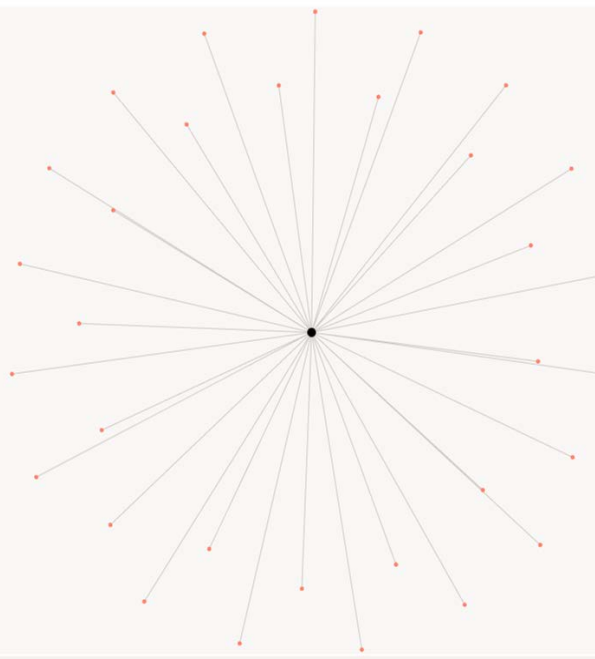
Outils :

Hyphe (crawler)

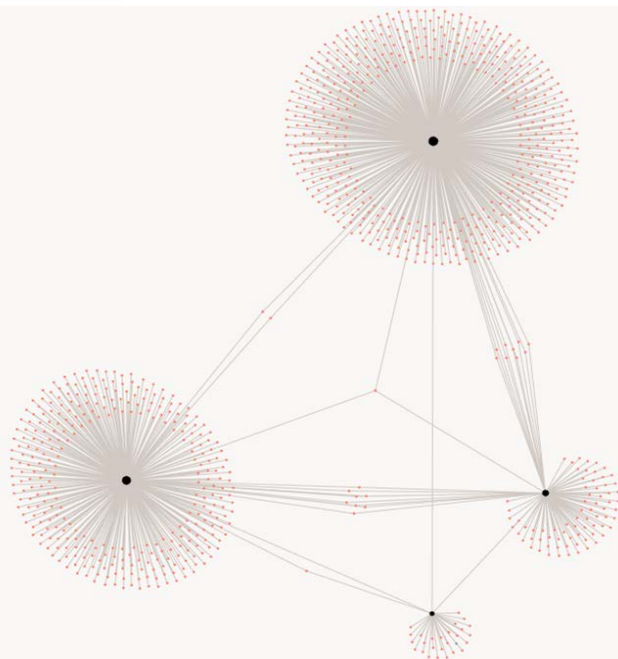
Google sheet

Gephi (visualisation)
Site web

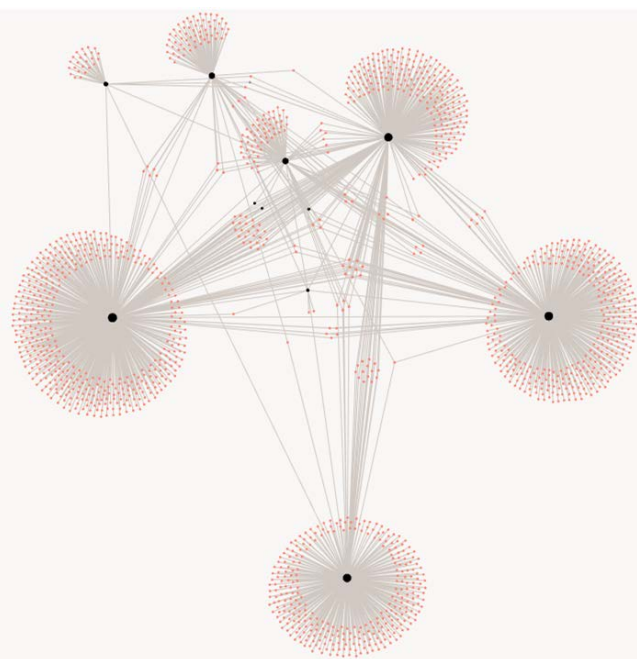
CORPUS EN CONSTRUCTION



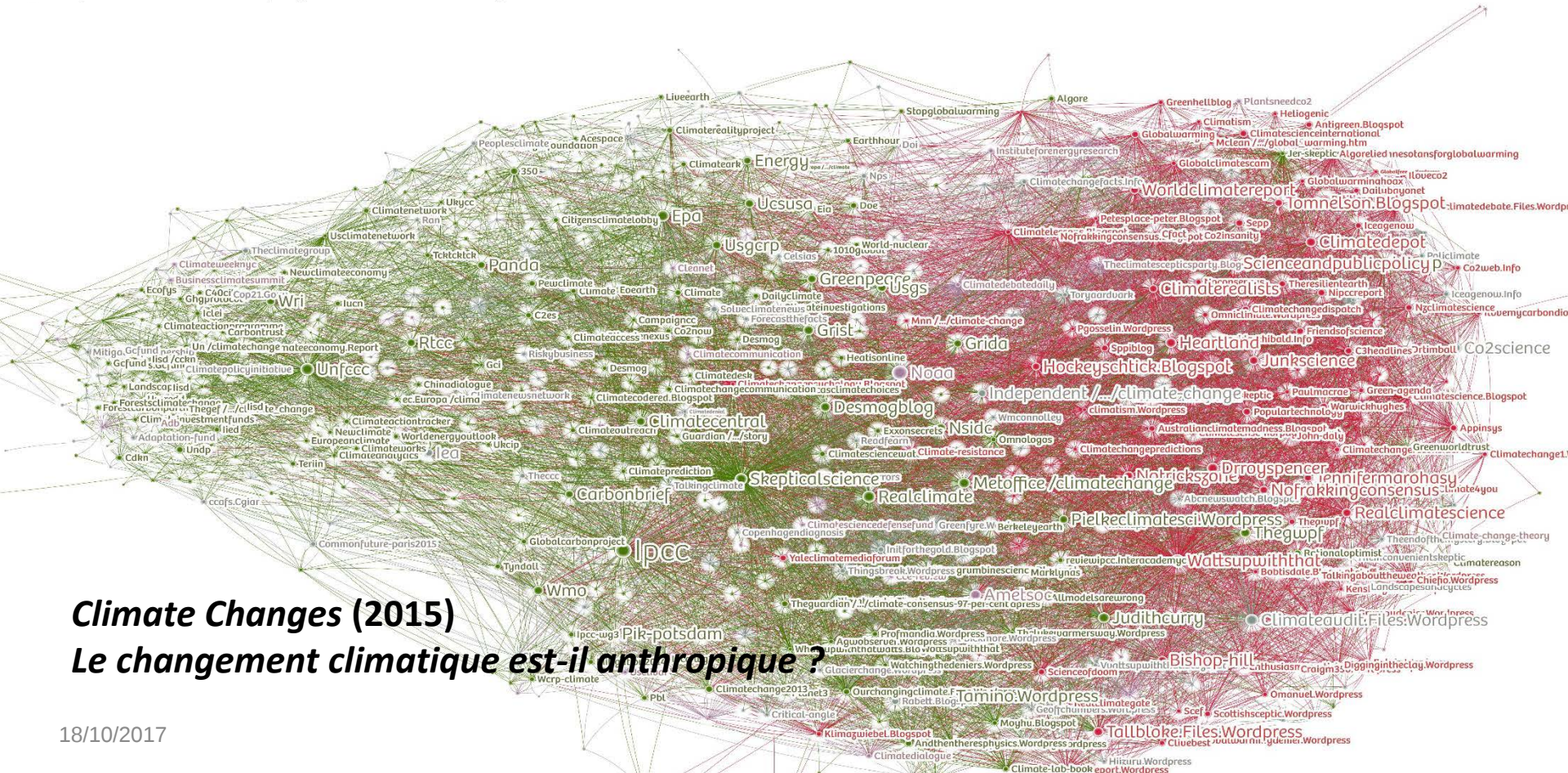
1



2

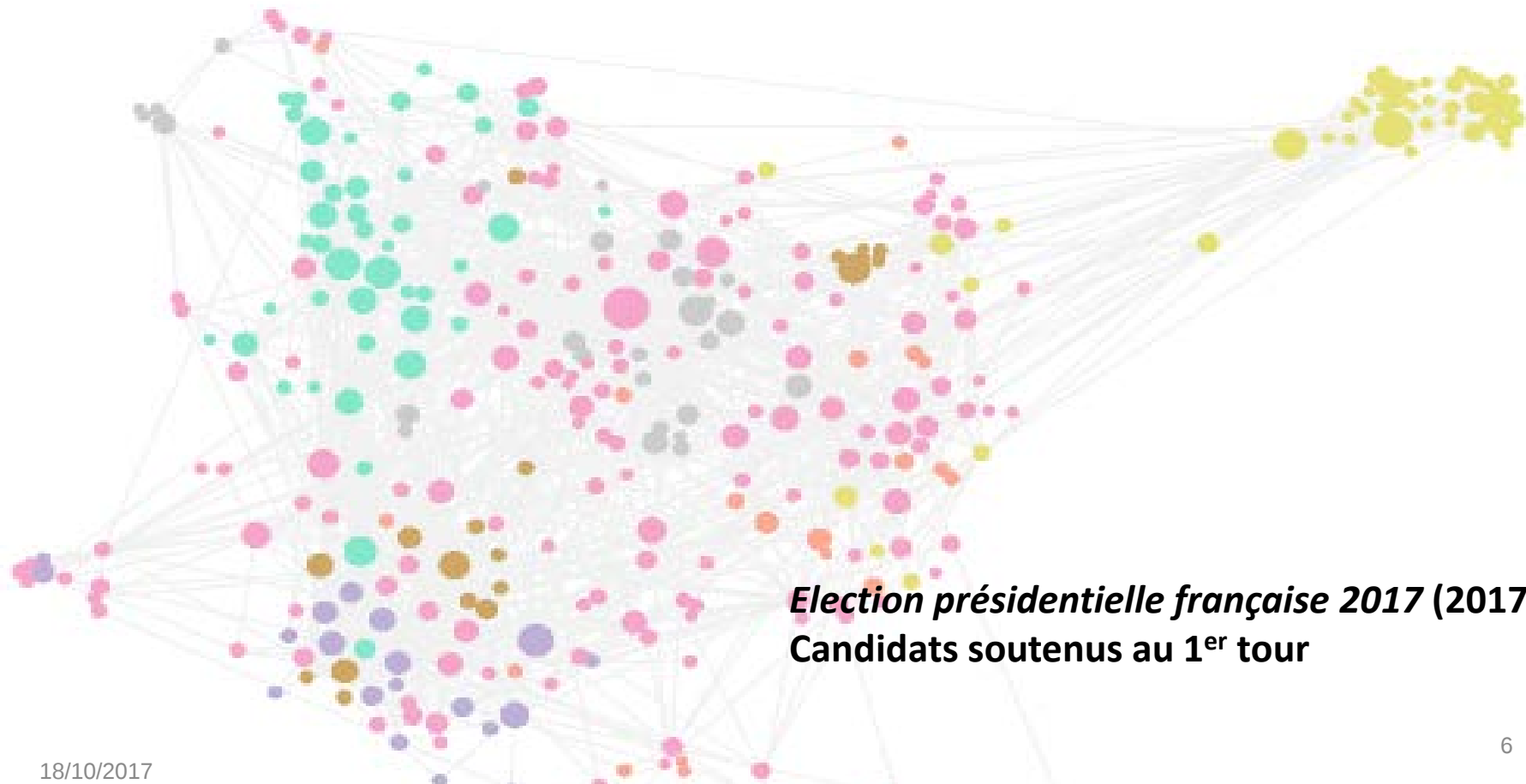


3



Climate Changes (2015)

Le changement climatique est-il anthropique ?



PROFILS ET COMPÉTENCES

Chercheurs

- définition et circonscription du corpus
- exploration et analyse

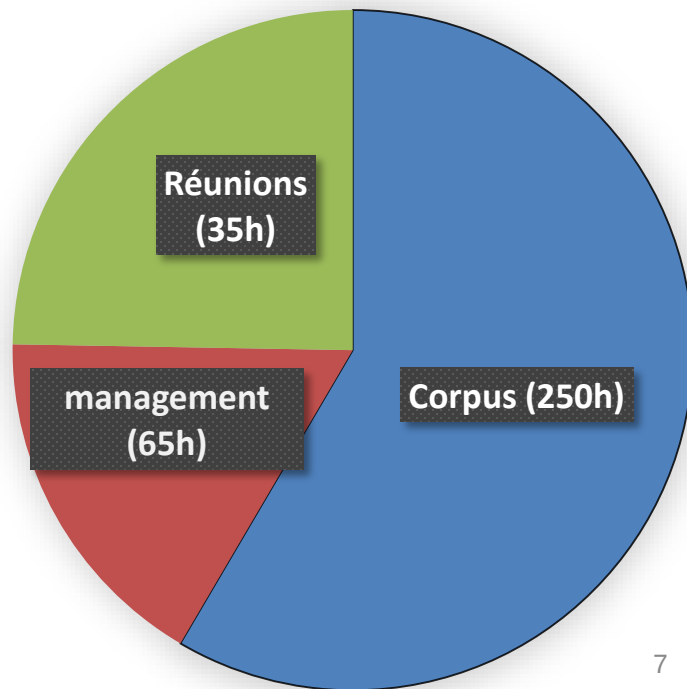
Ingénieurs spécialistes réseaux web et méthodes numériques

- méthodologie et outils (crawler, visualisation)
- design site web

Bibliothécaires

- constitution des données
- catégorisation et exploration
- coordination et outils (gestion des corpus)

Temps de travail par activité
Ex. : *Election présidentielle française 2017*
Total : 350h00



ÉLÉMENTS DE RÉFLEXION

Constats

- Travail à forte valeur ajoutée
- Confrontation de méthodes de travail différentes (linéaire, itératif)
- Mobilisation complexe des acteurs
- Donner de la profondeur et expliciter
- Analyser et prolonger la discussion scientifique

Expérimentation en cours

- Prochains corpus : quelles thématiques ? quelles modalités de travail ?
- Développer processus de travail collaboratif avec outils intégrés pour permettre à d'autres bibliothèques

EN SAVOIR PLUS

Corpus web

- Mathieu Jacomy, [*explorer les internets avec Hyphe*](#) (2017, METSEM)
- Corpus web à la bibliothèque : <http://corpusweb.sciencespo.fr>

Partenaires

- Médialab : <http://www.medialab.sciences-po.fr/fr/>
- Bibliothèque de Sciences Po : <http://www.sciencespo.fr/bibliotheque/>

Contact

- Cynthia Pedroja, chef de projet données de la recherche à la DRIS : cynthia.pedroja@sciencespo.fr

CORPUS WEB
ENTRE OUTIL DE RECHERCHE
ET OBJET DOCUMENTAIRE

18/10/2017