



NewsEye:
A Digital Investigator for Historical Newspapers

Antoine Doucet – Elisabeth Freyre

Sommaire

➤ NewsEye: présentation

- > Introduction et motivation
- > Objectifs
- > Organisation
- > Moyens et partenaires

➤ NewsEye: genèse

- > Collaborations bilatérales initiales
- > Appels CULT-COOP-09 et ANR MRSEI
- > Construction du consortium élargi
- > Préparation des réponses aux AAP

➤ Retours d'expérience

- > Sur le processus
- > Sur l'acceptation

➤ Bilan

- > Conclusion
- > Pour une interdisciplinarité réussie...



NewsEye: Présentation

*Projet soutenu par l'Union européenne
Horizon 2020*

NewsEye: présentation

Introduction



➤ Motivation

- > Presse historique: dizaines de millions de pages numérisées en Europe, nombreuses langues, nombreuses sources
- > Une mine d'or pour plusieurs disciplines SHS et le grand public
 - 2500 utilisateurs/jour sur ANNO; chercheurs et grand public (enseignants/élèves, généalogistes, activités de loisir)
- > Humanités numériques
 - Soutenir l'analyse qualitative par les données quantitatives,
 - Voire guider le qualitatif par le quantitatif

➤ Constat

- > Accès aux collections très difficile
 - Contenus de mauvaise qualité
 - Outils de recherche non adaptés à la presse ancienne
 - De nombreux documents sont « manqués » (requêtes non exhaustives, problèmes d'OCR et de structuration)...
 - Pas d'outils spécifiques pour l'analyse de la presse ancienne
 - Une quantité d'information inadaptée à une analyse par les experts

NewsEye: présentation

Objectifs

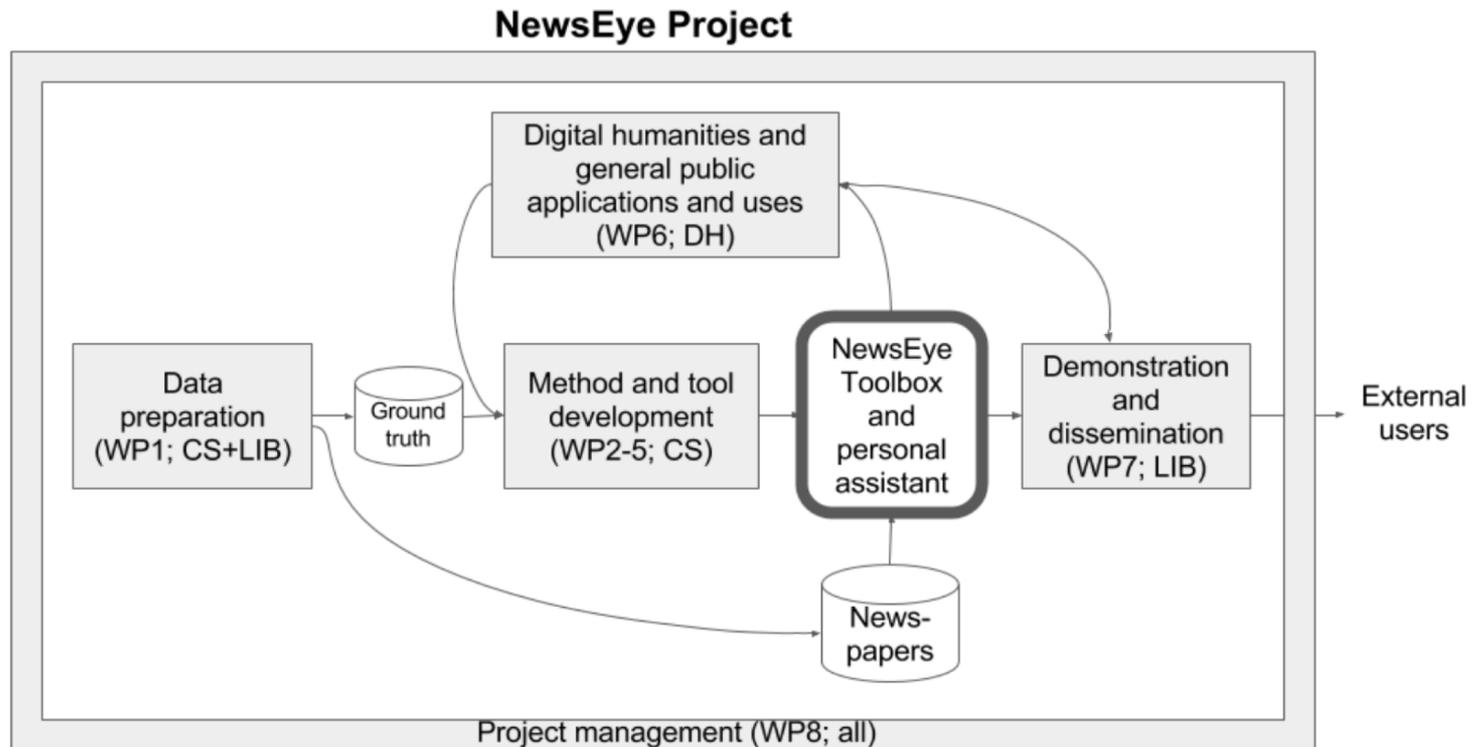


➤ Ambitions

- > Faciliter l'accès aux collections de presse ancienne
 - Réduire le bruit (OCR, structuration)
 - Enrichir les contenus sémantiquement
 - Inventer des méthodes robustes au bruit
 - Inventer des approches indépendantes des langues
- > Permettre une analyse quantitative...
 - qui soit justifiable et interprétable de façon qualitative
- > Impacter tous les utilisateurs potentiels de la presse historique
 - Grand public, enseignants, journalistes, bibliothécaires, étudiants, ...
 - Chercheurs en SHS (histoire, littérature, linguistique, géographie, arts du spectacle...)
- > **Toute langue, tout corpus, toute discipline**
 - immédiatement applicable à tout autre corpus (**approche durable**)
- > Contribuer à la recherche en SHS **et en informatique**

NewsEye: présentation

Organisation



NewsEye: présentation

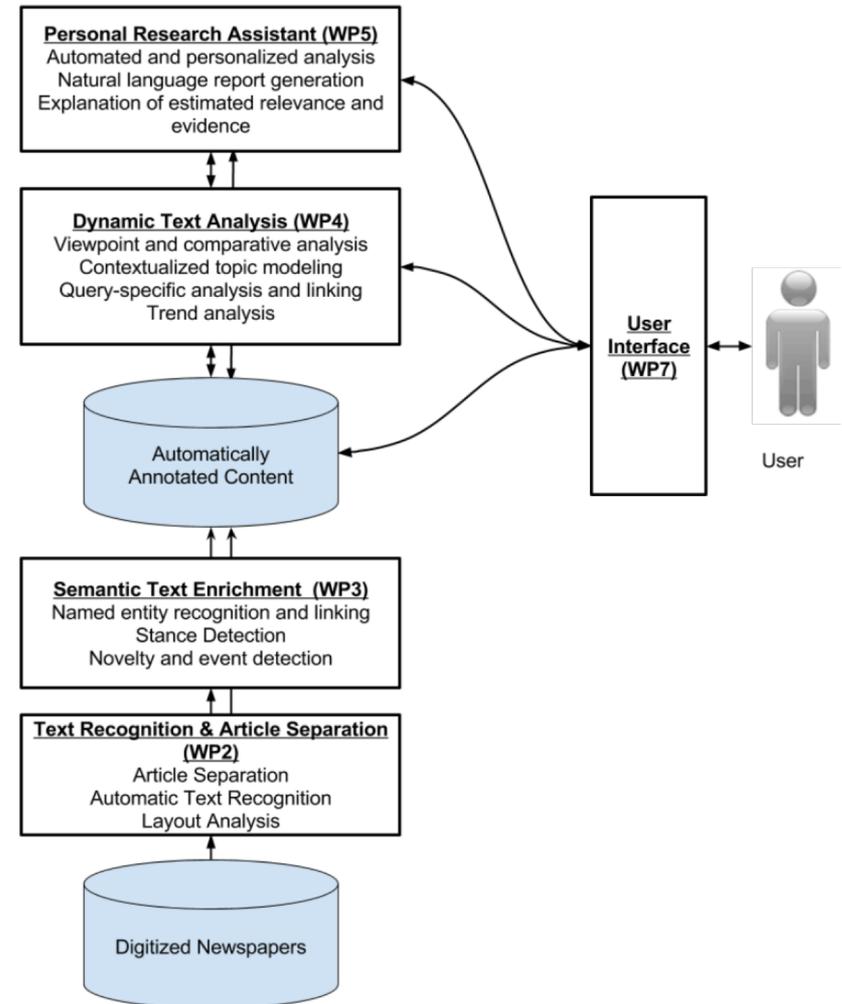
Organisation

➤ Au delà des livrables:

- > Production itérative et incrémentale
- > Accès en continu

➤ Pour les partenaires SHS

- > Accès et retour de pertinence en continu
- > Pas de phase d'attente



NewsEye: présentation

Moyens et partenaires

➤ Modalité

- > Financement: 2 998 565€
- > Durée: 3 ans (01/05/2018 - 30/04/2021)

➤ 4 pays, 5 langues principales, 3 types de partenaire (bibliothèque, SHS, informatique)

- | | | |
|--|-----|------------------|
| > Université de la Rochelle: | | Info |
| > Bibliothèque nationale d'Autriche | | Bibli |
| > Université d'Helsinki | | Bibli, SHS, Info |
| > Université d'Innsbruck | | SHS, Info |
| > Bibliothèque nationale de France | | Bibli |
| > Université de Rostock | | Info |
| > Université Paul Valéry Montpellier 3 | SHS | |
| > Université de Vienne | | SHS |

Sommaire

➤ NewsEye: présentation

- > Introduction et motivation
- > Objectifs
- > Organisation
- > Moyens et partenaires

➤ NewsEye: genèse

- > Collaborations bilatérales initiales
- > Appels CULT-COOP-09 et ANR MRSEI
- > Construction du consortium élargi
- > Préparation des réponses aux AAP

➤ Retours d'expérience

- > Sur le processus
- > Sur l'acceptation

➤ Bilan

- > Conclusion
- > Pour une interdisciplinarité réussie...

NewsEye: Genèse

Collaborations bilatérales initiales

➤ Travaux de recherche en TAL et RI

- > Analyse de contenus textuels, recherche d'information, veille épidémique multilingue sur la presse en ligne, ...
- > Spécificité: approches génériques, notamment alingues

➤ Avec l'université de Helsinki (Informatique)

- > Collaboration active depuis 2011
- > Co-encadrements et nombreuses co-publications
- > Nombreux échanges mais aucune réponse à « gros AAP » communs (aucun identifié)

➤ Avec l'université de Innsbruck (Informatique et histoire)

- > Rencontre en 2013, nombreux échecs sur AAP bilatéraux PRCI (analyse de la presse autrichienne pour l'histoire contemporaine)

➤ Avec la Bibliothèque nationale de France

- > Rencontre en 2015 (via Innsbruck !)
- > Projet AmeliOCR financé sur 2016-2017

➤ Interconnexion des collaborations déjà rêvée

- > Veille et analyse de texte (La Rochelle), analyse contrastive et résumé (Helsinki), numérisation, structuration et réseaux de bibliothèque (Innsbruck), histo-informatique (Innsbruck)
- > Plutôt comme la suite d'une ANR bilatérale...

➤ Opportunité 1 (2 février 2017):

- > H2020 Work programme part: SC6-CULT-COOP-2016-2017
- > CULT-COOP-09-2017: « *European cultural heritage, access and analysis for a richer interpretation of the past* »

➤ Opportunité 2 (juin 2016):

- > 3^{ème} appel ANR MRSEI

➤ Appel en 2 phases

- > 2 février 2017: 139 candidatures, 13 projets sélectionnés
- > 17 septembre 2017: 3 projets financés

NewsEye: genèse

Préparation de la réponse

- Réponse à l'AAP MRSEI: juin 2016
 - > Acceptation juillet 2017

- H2020 phase 1: septembre 2016 - février 2017
 - > Construction du consortium (du 1^{er} au 2^{ème} cercle)
 - > Réunion plénière à Helsinki (janvier 2016)
 - > Soumission le 2 février 2017

- H2020 phase 2: mai - septembre 2017
 - > Réponse phase 1: mai 2017
 - > Réunion plénière à Paris en juin 2017
 - > Été intense pour le coordinateur...
 - > Appui sur des consultants chevronnés et critiques
 - Retours constructifs et formateurs
 - > Soumission le 13 septembre 2017

- Méthodologie
 - > Travail collaboratif sur documents partagés
 - > Echanges réguliers en visio-conférence

Retour d'expérience

Retour d'expérience

Processus

➤ Bénéfices

- > Gain d'expertise énorme
 - grâce à la pratique
 - grâce aux partenaires
 - grâce aux consultants (financés par MRSEI)
- > Réseautage

➤ Points faibles

- > Taux de sélection
- > Travail à temps plein (dilettante impossible)

Retour d'expérience

Projet accepté !

➤ Bénéfices

- > Financement
 - et possibilités de co-financement
- > Visibilité (externe et interne)
- > Crédibilité élargie
- > Effet boule de neige

➤ Points faibles

- > Surcharge de travail

Bilan

Bilan

Foncez !

➔ On ne viendra (probablement) pas vous chercher

➔ Coordonner est le meilleur moyen de participer

- > Les partenaires sollicités ne diront pas non
- > Vous vous ferez connaître

➔ Faites-vous aider:

- > Sollicitez les compétences internes
- > Appuyez vous sur des partenaires aguerris
- > Aidez-vous : libérez du temps

➔ MRSEI:

- > Anticipation
- > Crédibilité
- > Renforce l'engagement mutuel des partenaires
- > Possibilité de pré-évaluation et d'accompagnement par des consultants

Bilan

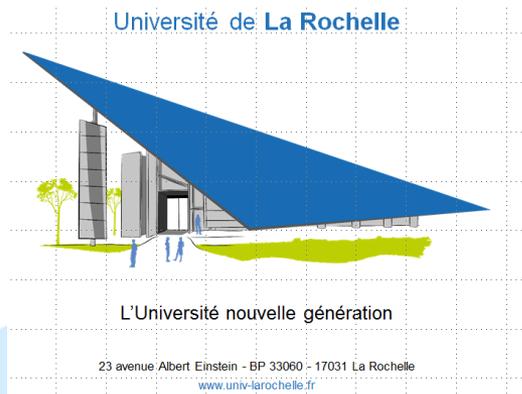
Interdisciplinarité

➤ Un projet interdisciplinaire doit bénéficier à toutes les disciplines

- > *Computer science* ≠ *Information Technology*
- > *Search* ≠ *Research*
- > Recherche en informatique ≠ mise en ligne de corpus
- > Clarifiez les besoins et attentes

➤ Excellence et impact doivent être interdisciplinaires

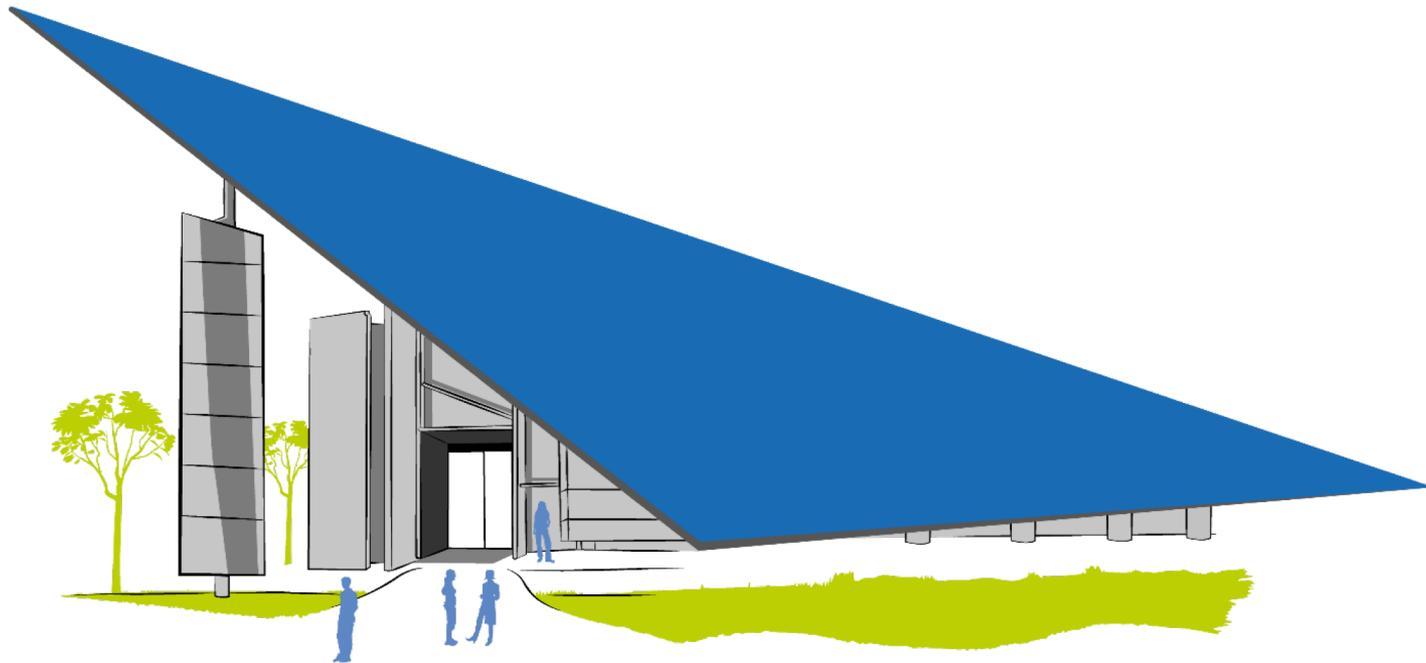
- > Impactez au delà du projet et de ses partenaires
- > (PS: bienvenus dans NewsEye)



{ BnF

Merci pour votre attention

Université de La Rochelle



L'Université nouvelle génération

23 avenue Albert Einstein - BP 33060 - 17031 La Rochelle

www.univ-larochelle.fr