

# Cartographier l'expertise des chercheurs et les accompagner vers la Science Ouverte :

## le cas de l'outil SOVisuHAL

**David Reymond\***, **Clara Galliano\*\***

\* MCF HDR, IMSIC (EA 7492), Université de Toulon, Aix Marseille Université, Toulon, France, david.reymond@univ-tln.fr

\*\* Doctorante, IMSIC (EA 7492), Université de Toulon, Aix Marseille Université, Toulon, France, clara-galliano@etud.univ-tln.fr



# Plan

1. Contexte, problématisation
  - a) Proposition : outil de cartographie (expertise explicite)
  - b) Domaines concernés (multidisciplinaire)
2. Accompagnement à la Science Ouverte
3. Résultats (prototype)
  - a) Technologies : TAL, Fouille, Expansion sémantique...
  - b) Fonctionnement de l'outil
  - c) Limites
  - d) Perspectives
4. Conclusion

# 1. Contexte et problématisation

- demande de l'équipe présidentielle de l'Université de Toulon en juillet 2019 (Reymond & Galliano, 2019)
- problématique : identifier les compétences d'expert dans les milieux académiques (Bouillot *et al.*, 2013; Zevio *et al.*, 2018)
- étude de l'existant : logiciels commerciaux, logiciels issus d'initiatives universitaires ou supportés par les éditeurs, supports « manuels »



## a. Proposition : outil de cartographie

- Expertise = ce qui est publié donc public et diffusé  
→ On collecte, traite, et représente tout cela, en aidant le chercheur à repérer où il en est...
- double intérêt : réponse à la demande de l'Université et incitation à la Science Ouverte
- Restriction: publications présentes sur HAL pour rallier les recommandations/directives européennes et étatiques à la reconstruction du “capital connaissance” de l'université
- le chercheur au centre du dispositif et acteur de sa lisibilité et visibilité doit valider et vérifier les données reconstituées : son image

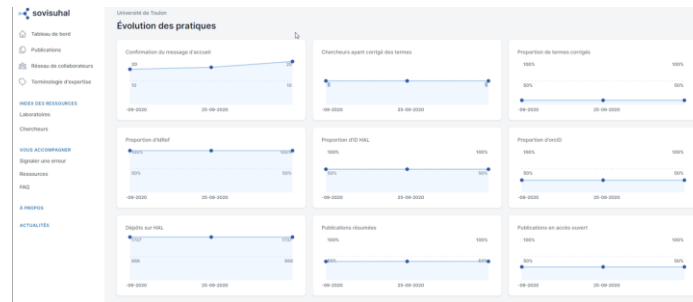
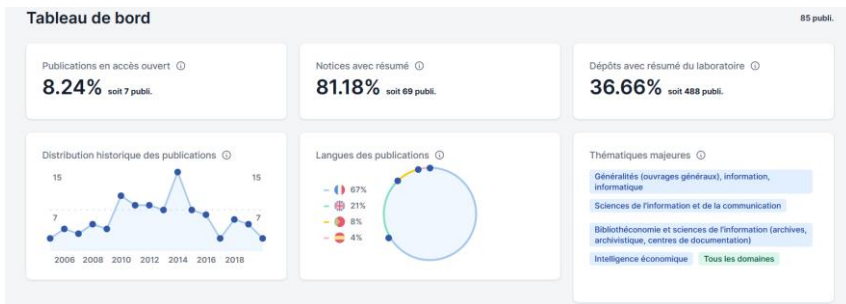
## b. Domaines concernés

- définition de l'expertise (Guittard & Welter, 2014 ; Bouillon, 2012 ; Shanteau, 1992)
- systèmes de recommandation d'experts : *expert finding system* et *expert profiling system*
- convoquer linguistique et informatique : le TAL pour rendre opérationnels des procédés de nettoyage et de filtrage nécessaires à l'extraction de connaissances et/ou d'entités nommées (Chakrabarti, 2018)
- l'utilisation de bases de connaissances ontologiques (Uren *et al.*, 2006) pour affiner les résultats ou encore développer une couche sémantique plutôt que statistique (Cifariello *et al.*, 2019) pour reconstituer des relations.

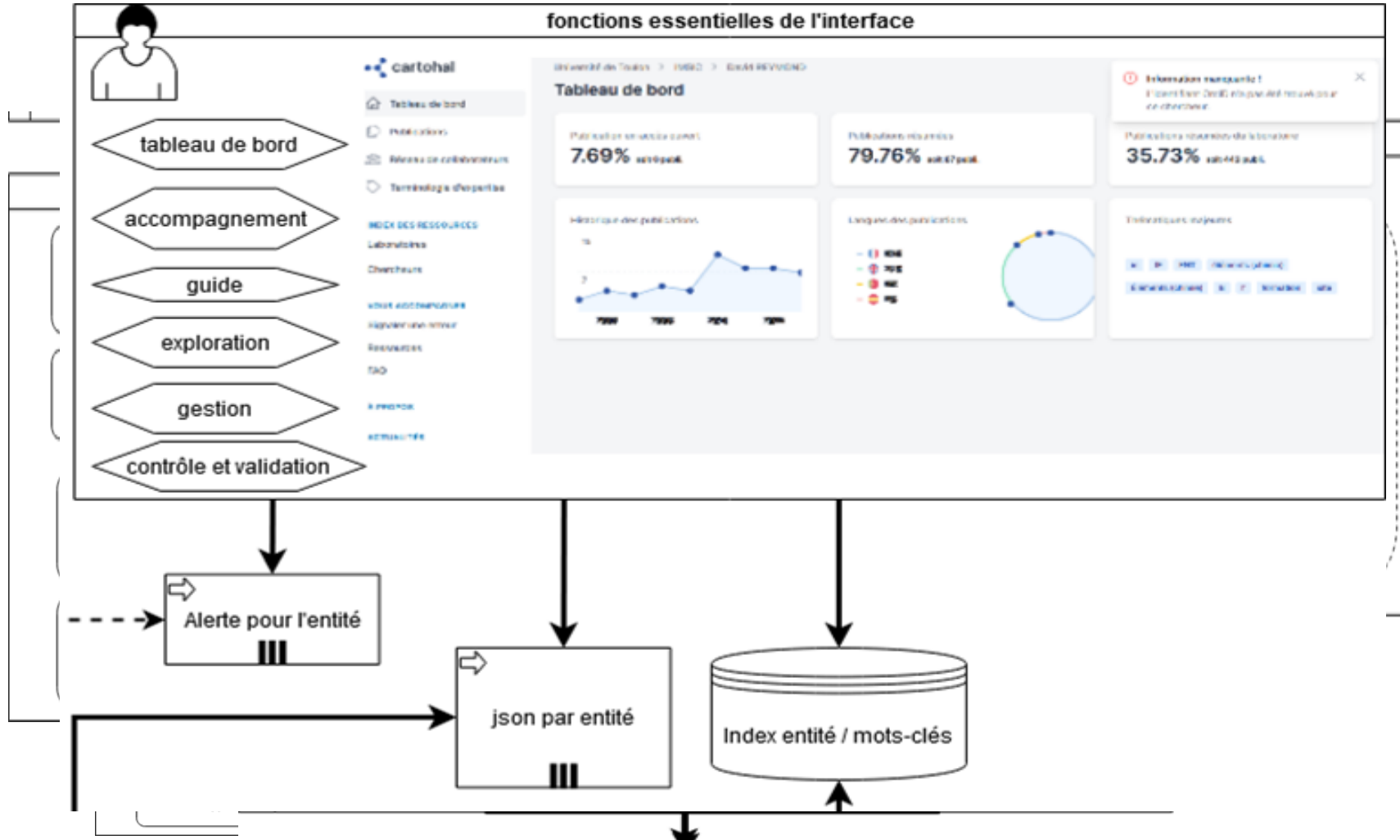
## 2. Accompagnement à la Science Ouverte

- *Loi pour une République Numérique* (2016)
- *Plan National pour la Science Ouverte* (juillet 2018)
- baromètre de la science ouverte (Jeangirard, 2019)
- contexte actuel : la diffusion des résultats de la recherche/des connaissances est largement encouragée par les politiques
- *Open Access* en France : présence de plateformes d'archives ouvertes, qu'elles soient nationales, thématiques, disciplinaires ou institutionnelles
- HAL c'est 2 359 288 de références bibliographiques : référentiel de base concernant le dépôt et la diffusion des connaissances

# De l'accompagnement instrumenté à la médiation à la science ouverte



# fonctions essentielles de l'interface





### 3. Résultats (prototype)

- un tableau de bord (laboratoire ou chercheur)
- une liste de publications identifiées présentées selon les trois vues : avec résumé et en OA, notices avec résumé; suggestion des **notices à compléter**
- un réseau des collaborateurs construit par rapport aux co-auteurs pour chaque publication identifiée
- un réseau des mots-clés repérable par couleur à partir des résumés existants de chaque entrée bibliographique
- espace documentation (tutoriel, aides, formations, ressources, FAQ)

# VOIR LA DÉMO



Université de Toulon > IMSIC

## Tableau de bord

1330 publ.

Publications en accès ouvert ⓘ

**21.35%** soit 284 publ.

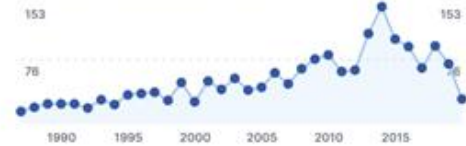
Notices avec résumé ⓘ

**36.17%** soit 481 publ.

Ratio ID HAL / chercheurs ⓘ

**60.47%** soit 26 chercheurs

Distribution historique des publications ⓘ



Langues des publications ⓘ

— 🇫🇷 82.78%  
— 🇪🇺 15.33%  
— 🌐 1.56%



Thématiques majeures ⓘ

Sciences de l'Homme et Société

Sciences de l'Homme et Société/Sciences de l'information et de la communication

Informatique et sciences de l'information

Sciences de l'information et de la communication **Tous les domaines**

- Tableau de bord
- Publications
- Réseau de collaborateurs
- Terminologie d'expertise

### INDEX DES RESSOURCES

Laboratoires

Chercheurs

### VOUS ACCOMPAGNER

Signaler une erreur

Ressources

FAQ

À PROPOS

ACTUALITÉS



## c. Limites

- Engagement humain à l'utilisation
- extraction des mots-clés auteurs complétée par l'extraction mots clés issus des résumés
- imperfections des traitements automatiques :
  - l'inconsistance des données initiales
  - application à des domaines disciplinaires variés
- limites des traitements réalisés :
  - typologie de document sur HAL non utilisée,
  - Non inclusion des citations,
  - Extraction des thématiques sur terminologie hors contexte

## d. Perspectives

- lancement officiel aux laboratoires pilotes et bêta-testeurs : IM2NP et IMSIC le lundi 21 septembre 2020
- Projection d'une V2 en janvier 2021 sur les autres laboratoires
- création d'indicateurs qualitatifs et quantitatifs
- intégration du dispositif dans l'ENT
- à terme : proposer l'outil à d'autres institutions
  - Moteurs de recommandations
  - Extracteurs « disciplinés »
  - Promotion d'outils de l'OA

## 4. Conclusion

- outil qui répond potentiellement aux objectifs
- chercheur : co-constructeur et expert des résultats produits par les dispositifs de traitement
- inscrire le dispositif dans une logique d'intelligence collective d'usage (Bourdoncle, 2010) au service de la science et de la gouvernance
- le dispositif proposé contourne ainsi les limites technologiques afférentes à l'extraction d'information, l'inconsistance initiale des données par une émergence collaborative (Levy, 2015)

# Quelques références

- Bouillot (F.), Gout (O.), Magnier (P.), Pénin (C.), Poncelet (P.), & Roche (M.), “Vers un outil de cartographie : Qui est l’expert ? “, *Démonstration du logiciel : 13e Conférence Francophone sur l’Extraction et la Gestion des Connaissances (EGC) 2013*, URL : <https://hal-lirmm.ccsd.cnrs.fr/lirmm-00798073>
- Bourdoncle (F.), “ L’intelligence collective d’usage“. In J.-M. Noyer & B. Juanals (Éd.), *Technologies de l’information et intelligences collectives*, Hermès Science Publications-Lavoisier, 2010
- Chakrabarti (S.), “Knowledge extraction and inference from text : Shallow, deep, and everything in between”, *The 41st international ACM SIGIR conference on research & development in information retrieval*, 2018, pp. 1399-1402.
- Cifariello (P.), Ferragina (P.), & Ponza (M.), “Wiser : A semantic approach for expert finding in academia based on entity linking”, *Information Systems*, 82, 2019, pp. 1-16.
- Explosion (A.), SpaCy-Industrial-strength natural language processing in python, 2017, URL: <https://spacy.io>
- Guittard, (C.), Welter (A.), Définition et sélection de l’expert IT : Approches cognitives et sociales. *Management & Avenir*, 67(1), 2014, pp. 101-121.  
<https://doi.org/10.3917/mav.067.0101>
- Jeangirard (E.), “Baromètre français de la science ouverte. Note Flash - Enseignement supérieur et Recherche (2016-...), 22, 2019, 2.
- Levy (P.), The emergence of reflexive collective intelligence, 2015, URL : <http://pierrelevyblog.com/tag/ieml/>
- Reymond (D.), Galliano (C.), “ Cartographie de l’expertise des chercheurs de l’université de Toulon. Etude préliminaire et prototype“, (rapport N° 1; 21 p.), Université de Toulon. 2019, URL : <https://hal.archives-ouvertes.fr/hal-02643329/document>
- Shanteau (J.), Competence in experts: The role of task characteristics. *Organizational behavior and human decision processes*, 53(2), 1992, pp. 252–266.
- Uren (V.), Cimiano (P.), Iria (J.), Handschuh (S.), Vargas-Vera (M.), Motta (E.), & Ciravegna (F.), “Semantic annotation for knowledge management : Requirements and a survey of the state of the art”, *Web Semantics: science, services and agents on the World Wide Web*, 4(1), 2006, pp. 14-28.
- Zevio (S.), Zargayouna (H.), Santini (G.), & Charnois (T.), « Vers une cartographie automatique des thématiques et profils d’experts associés à une conférence scientifique : 9 ans d’ateliers Recherche d’Information SEMantique (RISE) », *10ème Atelier Recherche d’Information SEMantique (RISE)*, 2018, URL : <https://hal.archives-ouvertes.fr/hal-02004675>

**Merci de votre attention**