

# Session d'information sur la Science Ouverte

---

Mardi 27 octobre



Co-funded by the  
Erasmus+ Programme  
of the European Union



**UBO**  
Université de Bretagne Occidentale

## Déroulé :

Introduction – SEA-EU & reSEArch-EU

Contexte et Enjeux de la Science Ouverte

Témoignages des ambassadeurs

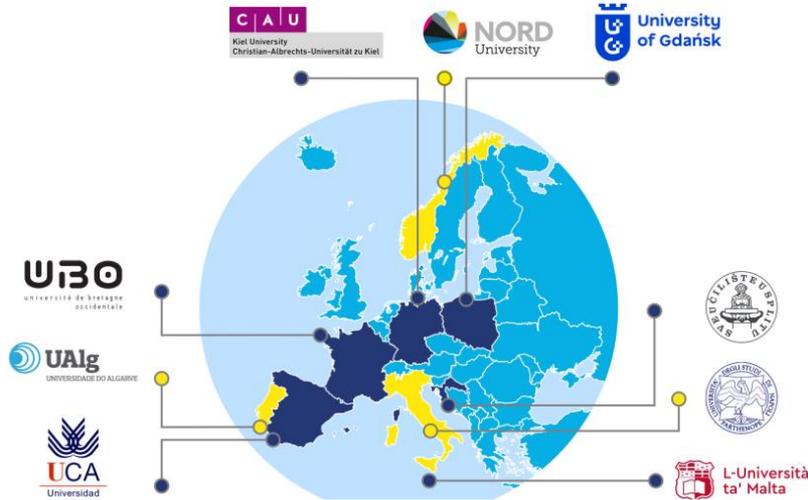
Présentations des workshops

# reSEArch-EU : soutenir la Recherche dans SEA-EU

---

Co-funded by the  
Erasmus+ Programme  
of the European Union





# SEA-EU: 9 universités, 1 Alliance, 3 objectifs

- Construire un **campus européen**, espace de formation et de mobilité pour tous les étudiants et les personnels des universités
  - Renforcer l'**identité européenne**
  - Répondre aux enjeux sociétaux de **développement durable**



# Objectifs de reSEArch-EU

## Science Ouverte

Plan de la recherche  
commun

Créer des liens avec  
d'autres alliance  
d'universités

Partage de capital  
humain et  
d'infrastructures



Résilience de la R&I

Alignement des pratiques  
d'innovation avec les  
besoins des parties  
prenantes

Ouvrir l'université vers la  
société

# Contexte et Enjeux de la Science Ouverte

---

# Définition

- La Science Ouverte concerne :
  - Le libre accès aux publications scientifiques
  - La gestion et le partage des données de la recherche
  - Les sciences participatives
- Focus Données de la recherche
- *« [...] des enregistrements factuels (chiffres, textes, images et sons), qui sont utilisés comme sources principales pour la recherche scientifique et sont généralement reconnus par la communauté scientifique comme nécessaires pour valider les résultats de la recherche. [...] Ces Principes et Lignes directrices portent essentiellement sur les données de la recherche sur support numérique exploitable sur ordinateur. »* OCDE, 2007. « Principes et lignes directrices de l'OCDE pour l'accès aux données de la recherche financée sur fonds publics ». Disponible sur : <http://www.oecd.org/fr/science/inno/38500823.pdf>

# Contexte : une injonction politique

- Programmes Horizon Europe et ANR
  - Le financement de la recherche publique sur projets impose la production du Plan de Gestion des Données (PGD/ DMP) en 2019
  - 10 juin 2022 : Conseil de l'UE adopte la réforme de l'évaluation de la recherche et la mise en œuvre de la SO
    - Recommande le libre accès immédiat à tous les résultats de la recherche
  - Voir aussi le récent Memorandum américain (août 2022)
- En France :
  - Loi pour une République Numérique de 2016 qui rend obligatoire l'ouverture des données de la recherche publique
  - Deuxième axe du Plan National pour la Science Ouverte (PNSO)
    - pour promouvoir politique des données **FAIR** (Trouvables, Accessibles, Interopérables, Réutilisables)

# Des arguments pour le partage

- Facilite à long terme la gestion de ses données
  - Valeur de preuve scientifique, revenir sur sa démarche, re-exploiter ses données
- Transparence et reproductibilité de la science dans contexte intégrité scientifique
  - Exigence des revues en lien avec la publication : vigilance
- Favorise collaborations, interdisciplinarité
  - Difficulté est davantage de trouver un langage commun autour des données
  - Potentiel de réutilisation?



**Le principe « aussi ouvert que possible, aussi fermé que nécessaire »**  
**Difficulté : déterminer le degré d'ouverture possible**

# Cadre juridique

## Obligation de publier

Données géographique et environnementales



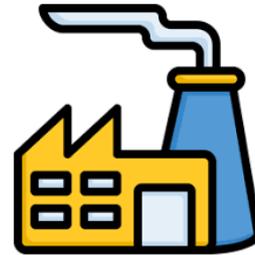
## Anonymiser

Données personnelles



## Protéger

Secret défense/ industriels, potentiel scientifique et technique de la nation



# Open Science / RGPD

- Règlement Général sur la Protection des Données / General Data Protection Regulation

# Principe

- Règlement européen transposé en droit français le 25 mai 2018
- Traitement : toute manipulation de données (y compris papier, copie, transfert, etc.)
- Objectif : protéger les droits fondamentaux des personnes (utilisation/réutilisation abusive, vie privée, etc.)

# Données personnelles

- Toute donnée liée à une personne physique
- Données directement / indirectement identifiantes
- e.g nom, prénom, date de naissance, etc.
- Données sensibles
- e.g code NIR (aka n° Sécurité Sociale), données de santé, opinions, orientations, etc.
- Principe de base : collecte et traitement doivent être licites

# Droits des personnes

- Consentement préalable à la collecte des données
- Consentement libre, spécifique et informée
- Cas des personnes vulnérables
- Droits d'opposition, d'accès, de rectification, de portabilité.  
Répondre sous 1 mois
- Ces droits sont limités dans certains cas (obligation légale, abus, etc.)

# Précautions d'usage en recherche (données personnelles)

- Séparez données primaires (collectées) et secondaires (résultantes)
- Anonymisez / pseudonymisez ASAP
- Protégez, sauvegardez, supervisez
- Détruisez ASAP

# RT vs ST

- Le Responsable de Traitement est tj responsable
- Le Sous-Traitant opère pour le compte du RT
- Contrat « RGPD » entre RT et ST
- Attention aux transferts hors UE ou pays adéquats
- Cloud Act, Privacy Shield, domaines juridiques mouvants

# Procédures de mise en conformité

- Inscription au registre des traitements de l'UBO
- Rédaction du protocole de recherche
- Méthodologies de Référence (MR)
- AIPD : analyse d'impact
- Comité éthique (éventuellement)
- Déclaration « CNIL » spécifique (éventuellement)

# Méthodologies de Référence

- 6 MR (cf. cnil.fr)
- La plus courante : MR-004 - Recherches n'impliquant pas la personne humaine, études et évaluations dans le domaine de la santé
- → concerne les traitements de données personnelles à des fins d'étude, d'évaluation ou de recherche n'impliquant pas la personne humaine

# Les risques

- Violation de données : fuite, perte, altération
  - Alerter et informer
  - Remédier
  - (re)sécuriser
- Recours des personnes et des tiers
- Croisement / désanonymisation

# Les sanctions

- Sanctions financières, pénales, professionnelles
- Image de marque
- Perte de crédibilité, impact sur l'avenir
- Injonction à stopper le traitement, à corriger
- Audits subséquents et charge de travail

# Témoignages des ambassadeurs

---

# Données Ouvertes /Open Data

## Axe de recherche UBO : Numérique Mathématiques

### Exemple : Domaine des Communications Numériques et du Traitement du signal

reSEArch-EU Session d'Information sur la Science Ouverte  
27 Septembre 2022

Roland GAUTIER

# Sommaire de la présentation

- Intérêts majeurs des Données Ouvertes dans les domaines Math-STIC 3
- Avantages / Inconvénients : Utilisateur vs Créateur 4
- Exemple 1 : Cas Utilisateur de données ouvertes 5  
Projet FEDER AN-DRO
- Exemple 2 : Cas Créateur de données ouvertes 7  
Projet H2020 MSCA-IF 2018 : UWB-IODA SF-PC

# Intérêts majeurs des Données Ouvertes dans les domaines Math-STIC

- Mise à disposition et utilisation en accès ouvert de :
  - Bases de données expérimentales et/ou Simulées : Signaux de com. , Signaux issus de capteurs diverses, Images, Sons, etc ...
  - Algorithmes et/ou programmes ;
  - Configurations de dispositifs expérimentaux (Testbed), systèmes et/ou montages ;
- Sur des plateformes de dépôt en ligne :
  - Comportant une très grosse capacité de stockage ;
  - Pérennes ;
  - Maintenues ;

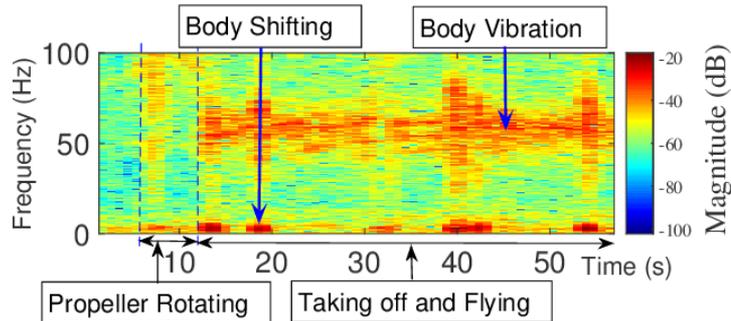
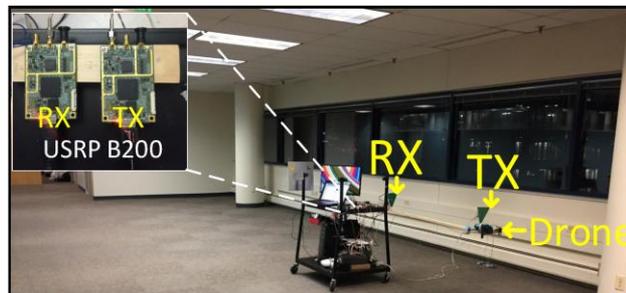
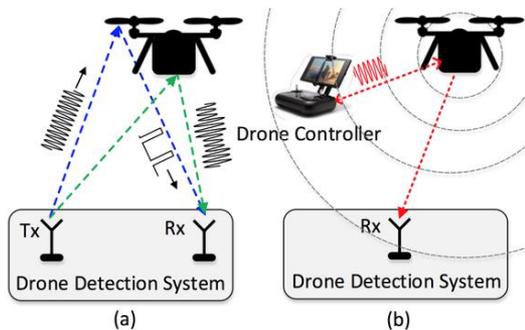
# Avantages / Inconvénients Utilisateur vs Créateur

	Avantages	Inconvénients / Devoirs / Verrous
Coté Utilisateur des données ouvertes	<ul style="list-style-type: none"><li>• Disposer de données expérimentales sans équipements spécifiques</li><li>• Réduction des coûts de la recherche, des délais de conception / traitements, des besoins en RH</li><li>• Accélérations des publications et de la dissémination des résultats</li><li>• Permettre des études comparatives sur des données de référence communes et partagées.</li></ul>	<ul style="list-style-type: none"><li>• Respect des conditions spécifiques d'utilisation de la licence imposée.</li></ul>
Coté Créateur des données ouvertes	<ul style="list-style-type: none"><li>• Citation des travaux et de la base de données potentielle (si réutilisée) ;</li><li>• Augmentation potentielle de l'acceptabilité d'une demande de financement.</li></ul>	<ul style="list-style-type: none"><li>• Coûts élevés des systèmes d'acquisition</li><li>• Chronophage</li><li>• Travail supplémentaire nécessaire pour le formatage et ajouts des méta data</li><li>• Risques de se faire devancer par la concurrence</li><li>• RH supplémentaires nécessaires</li><li>• Impossible si données sensibles ou sujettes au secret Indus. / Défense.</li><li>• De plus en plus imposé par les financeurs institutionnels.</li></ul>

# Exemple 1 : Cas Utilisateur de données ouvertes

## Projet FEDER AN-DRO

- AN-DRO : Analyse Numérique de signaux de DROne



“Matthan: Drone Presence Detection by Identifying Physical Signatures in the Drone's RF Communication” Phuc Nguyen, Anh-Hoang Truong, +3 authors Tam Vu Published in MobiSys 2017 DOI:10.1145/3081333.3081354

# Exemple 1 : Cas Utilisateur de données ouvertes

## Projet FEDER AN-DRO (suite)

### ORACLE RF Fingerprinting Dataset

<https://genesys-lab.org/oracle>

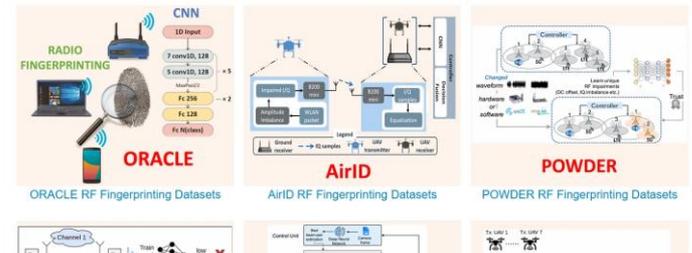
### AirID RF Fingerprinting Dataset

<https://www.genesys-lab.org/airid>

[GENESYS -Next GEneration NETworks and SYStems Lab](https://www.genesys-lab.org)  
Electrical and Computer Engineering Department - Northeastern  
University, Boston

Base de donnée de signaux échantillonnés pour validation de techniques d'authentification par empreinte radio permettant de détecter un émetteur particulier à partir d'un grand nombre de périphériques.

Format utilisé : [SigMF \(Signal Metadata Format\)](#)  
Licence « Creative Commons » : [CC-BY-SA-4.0 license](#)



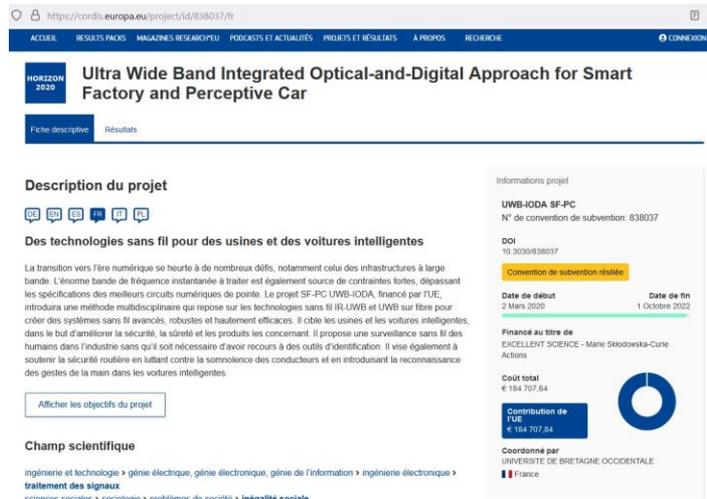
K. Sankhe, M. Belgiovine, F. Zhou, S. Riyaz, S. Ioannidis, and K. R. Chowdhury, "ORACLE: Optimized Radio Classification through Convolutional neural Networks," IEEE INFOCOM 2019, Paris, France, May, 2019.

K. Sankhe, M. Belgiovine, F. Zhou, L. Angioloni, F. Restuccia, S. D'Oro, T. Melodia, S. Ioannidis, and K. R. Chowdhury, "No Radio Left Behind: Radio Fingerprinting Through Deep Learning of Physical-Layer Hardware Impairments," IEEE Transactions on Cognitive Communications and Networking, Special Issue on Evolution of Cognitive Radio to AI-enabled Radio and Networks, 2019.

# Exemple 2 : Cas Créateur de données ouvertes

## Projet H2020 MSCA-IF 2018 : UWB-IODA SF-PC

- UWB-IODA SF-PC : Ultra Wide Band Integrated Optical-and-Digital Approach for Smart Factory and Perceptive Car
  - Programme H2020-EU.1.3. - EXCELLENT SCIENCE - Marie Skłodowska-Curie
  - Thème : MSCA-IF-2018 - Individual Fellowships



https://cordis.europa.eu/project/id/838037/fr

ACCEIL | RESEAUX PARTIS | MAGAZINES RESEAUX PARTIS | PODCASTS ET ACTUALITÉS | PROJETS ET RÉSULTATS | À PROPOS | RECHERCHER

### Ultra Wide Band Integrated Optical-and-Digital Approach for Smart Factory and Perceptive Car

Fiche descriptive | Résultats

#### Description du projet

🗣️ 📄 🇪🇺 🇫🇷 🇮🇹 🇱🇺

#### Des technologies sans fil pour des usines et des voitures intelligentes

La transition vers l'ère numérique se heurte à de nombreux défis, notamment celui des infrastructures à large bande. L'écosystème de fréquences instantanées à traiter est également source de contraintes fortes, dépassant les spécifications des meilleurs circuits numériques de pointe. Le projet SF-PC UWB-IODA, financé par l'UE, introduit une méthode multidisciplinaire qui repose sur les technologies sans fil IR-UWB et UWB sur fibre pour créer des systèmes sans fil avancés, robustes et hautement efficaces. Il cible les usines et les voitures intelligentes, dans le but d'améliorer la sécurité, la sûreté et les produits les concernant. Il propose une surveillance sans fil des humains dans l'industrie sans qu'il soit nécessaire d'avoir recours à des outils d'identification. Il vise également à soutenir la sécurité routière en luttant contre la somnolence des conducteurs et en introduisant la reconnaissance des gestes de la main dans les voitures intelligentes.

[Afficher les objectifs du projet](#)

#### Champ scientifique

ingénierie et technologie > génie électrique, génie électronique, génie de l'information > ingénierie électronique > traitement des signaux  
sciences sociales > sociologie > problèmes de société > inégalité sociale

#### Informations projet

**UWB-IODA SF-PC**  
N° de convention de subvention: 838037

DOI  
10.3030/838037

Convention de subvention révisée

Date de début  
2 Mars 2020

Date de fin  
1 Octobre 2022

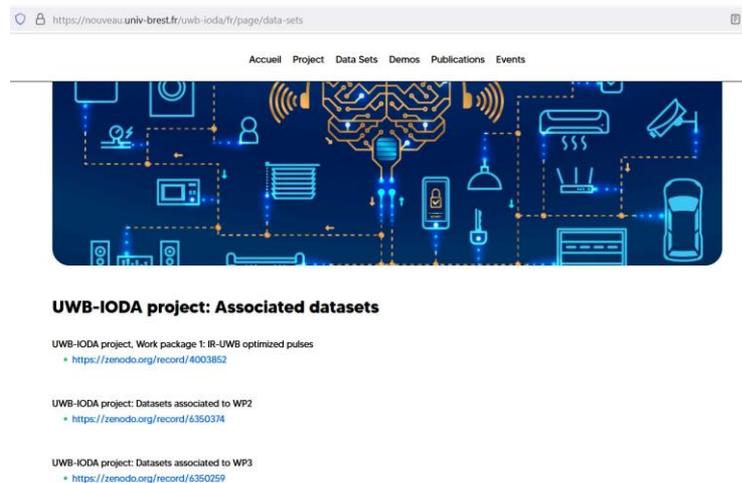
Financé au titre de  
EXCELLENT SCIENCE - Marie Skłodowska-Curie  
Actions

Coût total  
€ 154 707,84

Contribution de l'UE  
€ 164 707,84

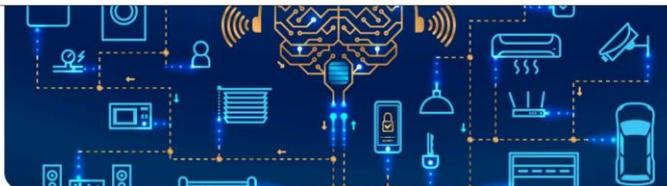
Coordonné par  
UNIVERSITÉ DE BRETAGNE OCCIDENTALE  
France

<https://cordis.europa.eu/project/id/838037/fr>



https://nouveau.univ-brest.fr/uwb-ioda/fr/page/data-sets

Accueil | Project | Data Sets | Demos | Publications | Events



### UWB-IODA project: Associated datasets

UWB-IODA project, Work package 1: IR-UWB optimized pulses  
• <https://zenodo.org/record/4003852>

UWB-IODA project: Datasets associated to WP2  
• <https://zenodo.org/record/6350374>

UWB-IODA project: Datasets associated to WP3  
• <https://zenodo.org/record/6350259>

<https://nouveau.univ-brest.fr/uwb-ioda/fr>

# Exemple 2 : Cas Créateur de données ouvertes

## Projet H2020 MSCA-IF 2018 : UWB-IODA SF-PC (suite)

**WB-IODA project, Work package 1: IR-UWB optimized pulses**

<https://zenodo.org/record/4003852>

**UWB-IODA project: Datasets associated to WP2**

<https://zenodo.org/record/6350374>

**UWB-IODA project: Datasets associated to WP3**

<https://zenodo.org/record/6350259>

**Licence « Creative Commons »:**

**CC-BY-4.0 license**

Référencé dans : [OpenAIRE Explore](https://explore.openaire.eu/)

<https://explore.openaire.eu/>

(Service des projets H2020 OpenAIRE et OpenAIREplus) portés en outre par le CERN

The screenshot shows a Zenodo record page for the dataset 'UWB-IODA project, Work package 1: IR-UWB optimized pulses'. The page includes the Zenodo logo, a search bar, and navigation links for 'Upload' and 'Communities'. The record is dated August 27, 2020, and has 51 views and 2 downloads. The authors listed are Radoi, Emanuel; Khan, Faheem; Faheem Khan; and Stéphane Azou, Roua Youssief, Pascal Morel. The project is associated with OpenAIRE. The record has received funding from the European Union's Horizon 2020 research and innovation programme. The data files are listed as follows:

Name	Size	Download
UWB_bpslimes_monopulse_optimization_ECC_mask.mat	8.5 kB	Download
UWB_bpslimes_monopulse_optimization_FCC_mask.mat	6.3 kB	Download

Dépôt sur : [Zenodo](https://zenodo.org/)

Dépôt international générique d'OpenAIRE, hébergé au CERN et entièrement compatible avec les exigences d'OpenAIRE.

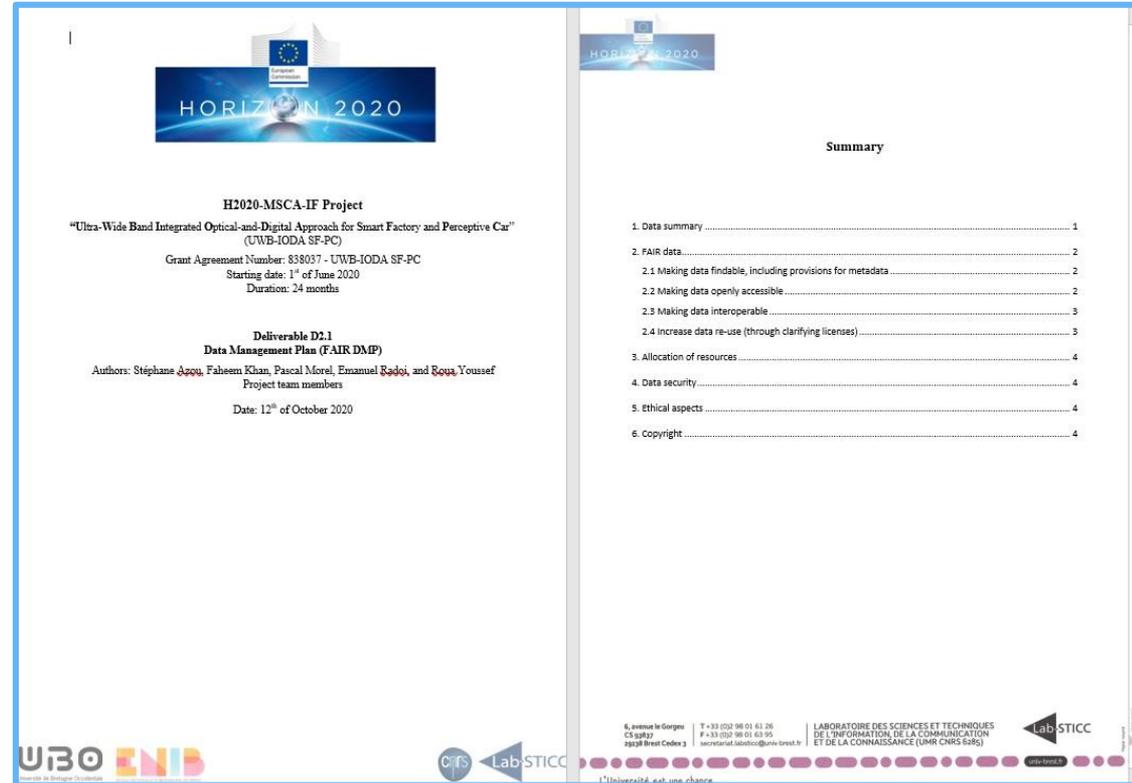
Obtention d'un DOI pour les jeux de données :

Exemple: DOI:10.5281/zenodo.4003852 pour WB-IODA project, Work package 1: IR-UWB optimized pulses

# Exemple 2 : Cas Créateur de données ouvertes

## Projet H2020 MSCA-IF 2018 : UWB-IODA SF-PC (suite)

- Obligation réglementaire pour H2020 de rédiger parmi les livrables un Plan de Gestion des Données ou Data Management Plan (DMP) après le démarrage du projet ;
- Compatible avec ou respectant le principe FAIR :
  - Findable ;
  - Accessible ;
  - Interoperable ;
  - Reusable.



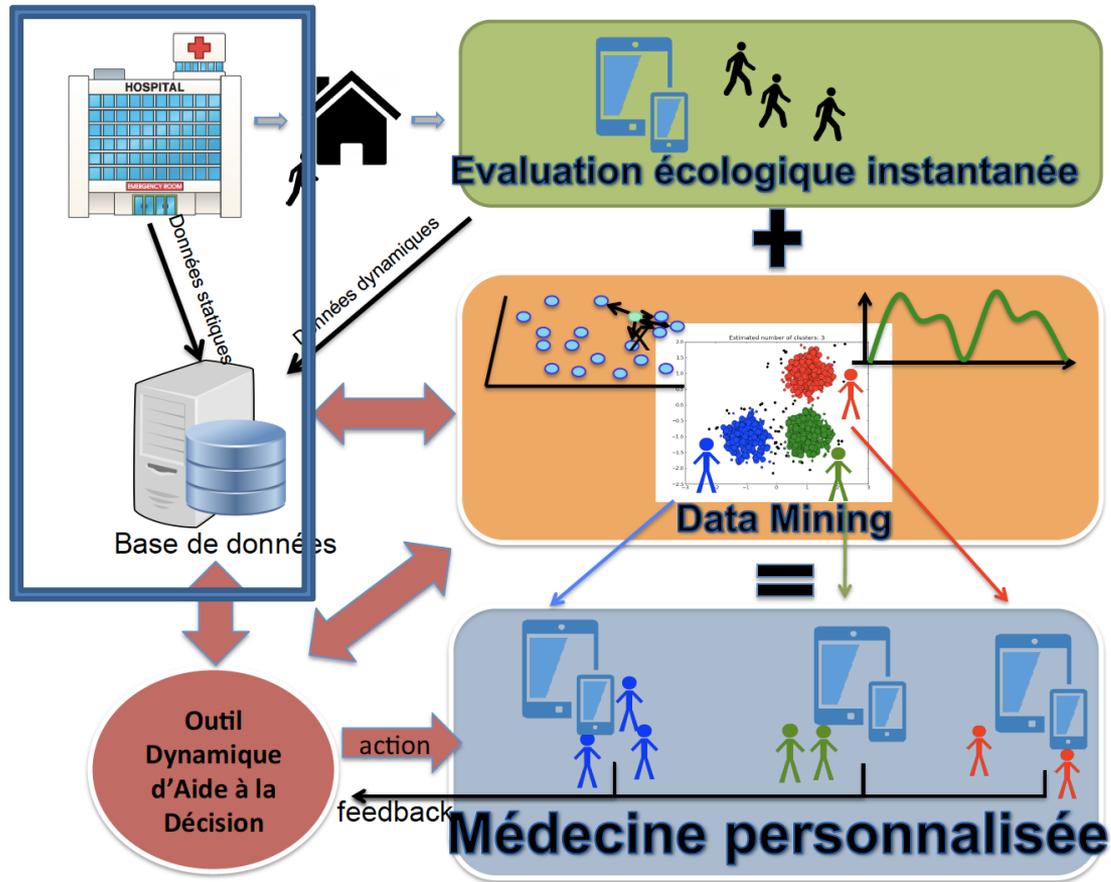
**Merci pour votre attention !**

# Partie santé

## Enjeux de la donnée en santé

# Place de la donnée en santé

- Appuyer le professionnel de santé dans un contexte clinique qui se complexifie.
- Produire de nouvelles connaissances
- Améliorer la compréhension et la transparence du système de santé.
- Prendre en charge les patients le plus tôt possible.
- Prédire les trajectoires individuelles des patients et améliorer les actions de prévention.



# Projet EURECA

- Base de données de 2800 patients admis aux urgences et 700 variables (France + Espagne)
- 681 patients atteignaient le seuil minimal de 70% (23 variables)
- Analyse visant à établir des clusters de risque

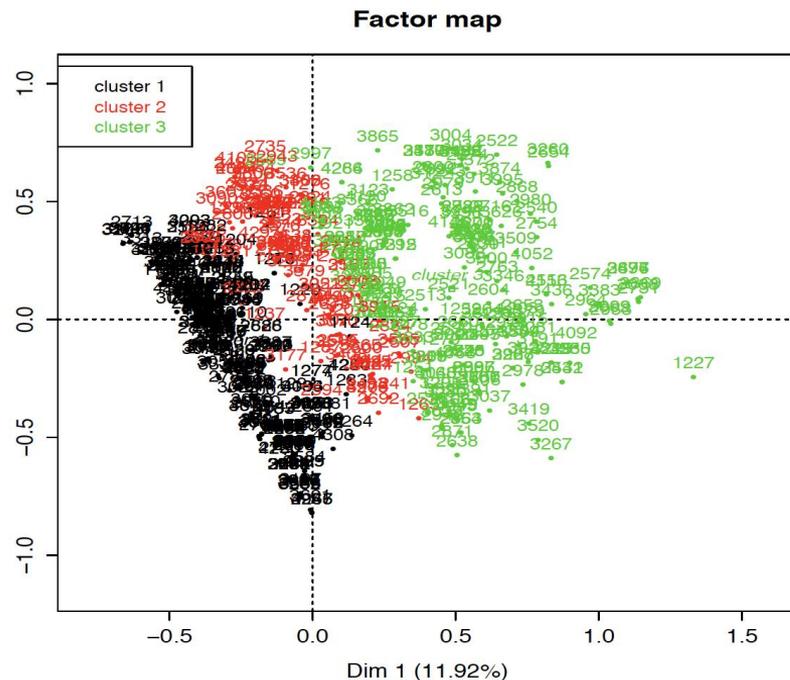
# Analyse initiale

- Une approche non supervisée
- Dossier médicaux informatisée (Questionnaires structurés)
- Patients admis aux urgences après un geste suicidaire
- La base de données originale contenait 2802 patients et 263 variables
- Jeu final de données concernant 681 patients et 20 variables

# Méthode

- Multiple Correspondence Analysis (MCA) : réduction du nombre de variables
- Hierarchical Clustering on Principal Components (HCPC)
  - Représenter les patients dans un espace qui permettrait de mettre en évidence des groupes

# Analyse des correspondances multiples (ACM)



# Perspectives de MEmind



*Berrouiguet, et al, accepted*

# Enjeux des guichets uniques

- Soutenir les responsables de données dans la collecte, la consolidation et la mise en valeur de leur patrimoine
- Valorisation et rétribution du travail et de l'expertise du responsable des données
- Garantir à la société civile le respect des droits des citoyens
- Simplifier l'accès à la donnée
- Innover aux côtés des acteurs de la recherche et de l'industrie

# DONNÉES DE SANTÉ



Source : HDH

**Merci pour votre attention !**

# Sciences ouvertes et Lettres\*: Retour d'expériences

Liana Ermakova

liana.ermakova@univ-brest.fr

Disclaimer: L'ouverture des données est en cours...

# Pourquoi partager des ressources linguistiques?

Pour faciliter la recherche

Pour produire des résultats comparables

Chaque publication qui utilise les données doit les citer

# Impact de l'ouverture des données sur la nombre de citations



Ellen Voorhees

[NIST](#)

Verified email at nist.gov

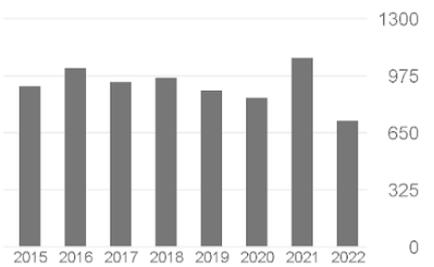
[information retrieval](#)

FOLLOW

TITLE	CITED BY	YEAR
<b>Overview of TREC 2002.</b> EM Voorhees, D Harman Trec	1961*	2002
<b>Overview of the sixth text retrieval conference (TREC-6)</b> EM Voorhees, D Harman Information Processing & Management 36 (1), 3-35	1454*	2000
<b>Query expansion using lexical-semantic relations</b> EM Voorhees SIGIR'94, 61-69	1413	1994
<b>The trec-8 question answering track report.</b> EM Voorhees Trec 99, 77-82	1266	1999
<b>TREC: Experiment and evaluation in information retrieval</b> EM Voorhees, DK Harman MIT press	1125	2005

Cited by [VIEW ALL](#)

	All	Since 2017
Citations	23514	5442
h-index	63	39
i10-index	121	69



Public access [VIEW ALL](#)

0 articles	9 articles
not available	available

Based on funding mandates





## Paolo Rosso

FOLLOW

Full Professor, Computer Science, [Universitat Politècnica de València](#)  
Verified email at dsic.upv.es - [Homepage](#)

[Natural Language Processing](#) [Fake News detection](#) [Hate Speech detection](#) [Irony detection](#)  
[Artificial Intelligence](#)

GET MY OWN PROFILE

TITLE	CITED BY	YEAR
<a href="#">Overview of the 3rd international competition on plagiarism detection</a>	612 *	2011
<a href="#">Semeval-2019 task 5: Multilingual detection of hate speech against immigrants and women in twitter</a>	567	2019
<a href="#">A multidimensional approach for detecting irony in twitter</a>	488	2013

M. Rothbart, A. Eiselt, J. A. Barrán-Cedeño, B. Stein, P. Rosso

CEUR workshop proceedings 1177

twitter

V. Basile, C. Bosco, E. Fersini, D. Nozza, V. Patti, FMR Pardo, P. Rosso, ...

Proceedings of the 13th international workshop on semantic evaluation, 54-63

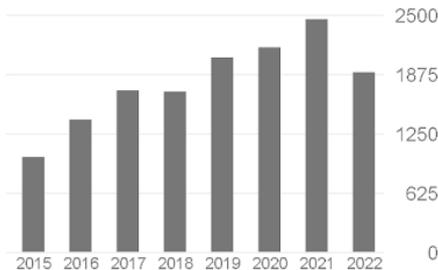
A. Reyes, P. Rosso, T. Veale

Language resources and evaluation 47 (1), 239-268

Cited by

VIEW ALL

	All	Since 2017
Citations	18415	12010
h-index	67	56
i10-index	307	217





## Gerhard Weikum

Professor of Computer Science, [Max Planck Institute for Informatics](#), Saarland Informatics Campus

Verified email at mpi-inf.mpg.de

[Databases](#) [Data Mining](#) [Information Retrieval](#) [Web Data Management](#) [Knowledge Bases](#)



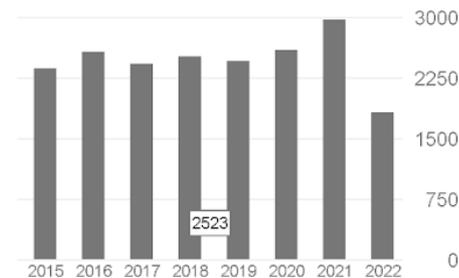
[GET MY OWN PROFILE](#)

Cited by

[VIEW ALL](#)

TITLE	CITED BY	YEAR
<a href="#">Yago: a core of semantic knowledge</a> FM Suchanek, G Kasneci, G Weikum Proceedings of the 16th international conference on World Wide Web, 697-706	4665	2007
<a href="#">YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia</a> J Hoffart, FM Suchanek, K Berberich, G Weikum Artificial intelligence 194, 28-61	1429	2013
<a href="#">The LRU-K page replacement algorithm for database disk buffering</a> EJ O'neil, PE O'neil, G Weikum Acm Sigmod Record 22 (2), 297-306	1413	1993
<a href="#">Robust disambiguation of named entities in text</a> J Hoffart, MA Yosef, I Bordino, H Fürstenau, M Pinkal, M Spaniol, ...	1161	2011

	All	Since 2017
Citations	41259	14835
h-index	97	52
i10-index	426	211



## WordNet: a lexical database for English

GA Miller - Communications of the ACM, 1995 - dl.acm.org

... in **WordNet**. Inflectional morphology for each syntactic category is accommodated by the interface to the **WordNet** ... dancing” are all distinct words in **WordNet**. A much larger variety of ...

☆ Save  Cite Cited by 17783 Related articles All 27 versions Autres liens 

# Retour d'expérience

- L'ouverture des données est en cours
- Valorisation des données dans les campagnes d'évaluation des algorithmes (CLEF)
- Financement obtenu pour la création des données
  - projet JOKER: Automatic Wordplay Analysis
    - SEA-EU search
    - MSHB
  - projet SimpleText: Automatic Text Simplification
    - GDR MADICS <https://www.madics.fr/>
    - ANR JCJC – offre d'une thèse financée !
- Défis majeur: le droit d'auteur

# CLEF 2022 Conference and Labs of the Evaluation Forum

*Information Access Evaluation meets Multilinguality, Multimodality, and Visualization*

**5 - 8 September 2022, Bologna - Italy**

- Evaluation of information access systems through experimentation on shared tasks
  - Data
  - Evaluation metrics
- a set of labs and workshops designed to test different aspects of mono and cross-language Information retrieval systems
- <https://clef2022.clef-initiative.eu/index.php>



# Project JOKER\*: Automatic Wordplay Analysis



Site web: <http://www.joker-project.com/>

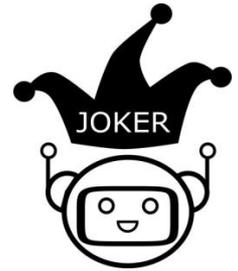
Email : [contact@joker-project.com](mailto:contact@joker-project.com)



\* This project has received a government grant managed by the National Research Agency under the program "Investissements d'avenir" with the Reference ANR-19-GURE-0001& the MSHB funding

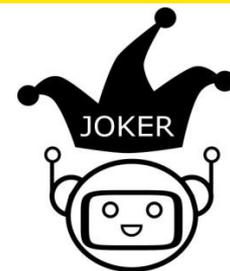


# Parallel dataset



- Manually collected existing translations (>2,000): video games, literature, advertising, ... ✓
- Our translations:
  - Puns from the Internet (master students in translation interns) ✓
  - Translation contest (> 5,000) ✓
    - English -> French
    - SemEval-2017 Task 7: Detection and Interpretation of English Puns (Miller et al., 2017)

# Tasks 2023



- Task 1: Pun detection in English, French and Spanish
- Task 2: Pun location and interpretation in English and French (and Spanish?)
- Task 3: Pun translation from English to French and to Spanish
- Open task (e.g. wordplay generation)

# Les exceptions au droit d'auteur - article L122-5 (SATT Ouest Valorisation)

## L122-5, 3°, a – exception de courte citation:

- la reproduction ne doit pas porter atteinte au droit moral de l'auteur ;
- la citation doit mentionner la source, l'auteur et le cas échéant, l'éditeur ;
- la citation doit être courte, au regard notamment de la longueur de l'œuvre citée, et de l'importance qualitative de l'extrait ;
- l'utilisation doit être justifiée par le caractère pédagogique ou scientifique.

## L122-5, 3°, e – exception à des fins d'enseignement et de recherche:

- la reproduction doit être faite à des fins d'enseignement et de recherche;
- l'auteur doit être rémunéré ;
- la reproduction doit être faite à destination d'un public composé majoritairement d'élèves, d'étudiants, d'enseignants et de chercheurs.



# Project SimpleText\*: Automatic Text Simplification

TURN IT SIMPLE



Site web: <https://simpletext-project.com/>

Email : [contact@simpletext-project.com](mailto:contact@simpletext-project.com)

\*ANR JCJC'2022





TURN IT SIMPLE

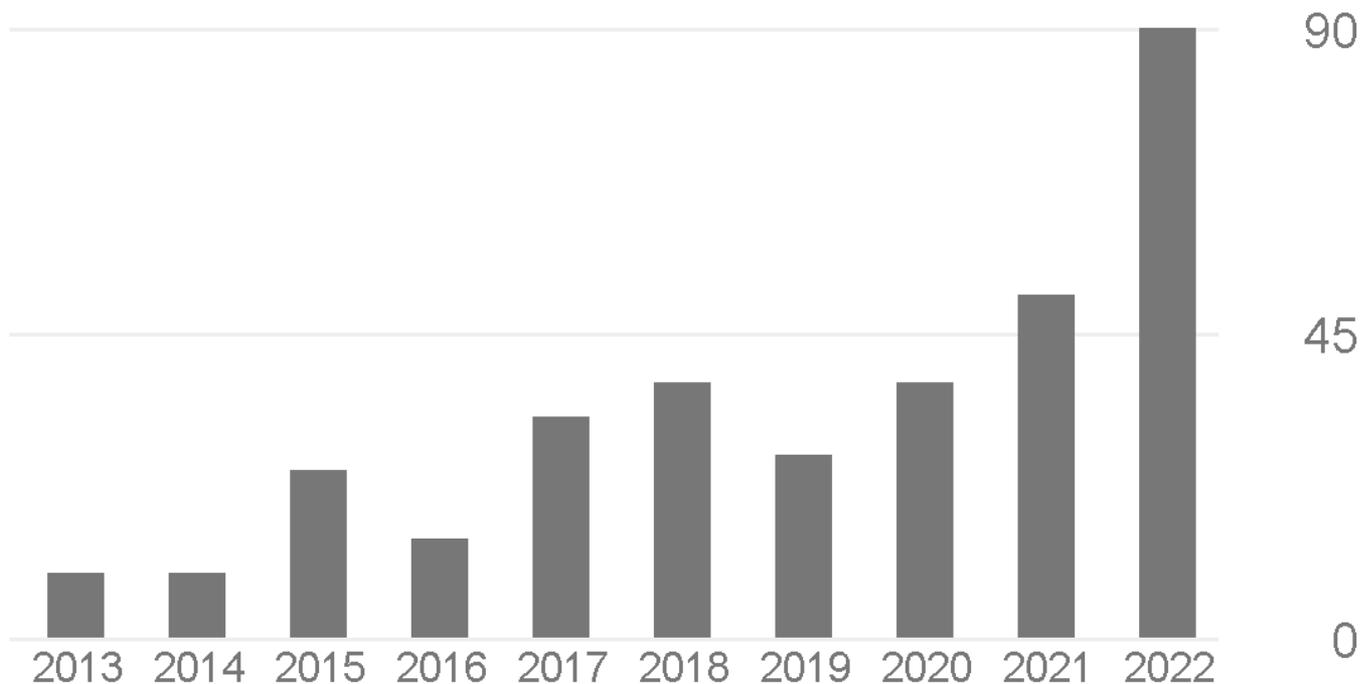


# Shared tasks

- **TASK 1** *What is in (or out)?*
  - Select passages to include in a simplified summary, given a query
- **TASK 2** *What is unclear?*
  - Given a passage and a query, search for terms/concepts that are required to be explained for understanding this passage (definitions, context, applications,..)
- **TASK 3** *Rewrite this!*
  - Given a query, simplify passages from scientific abstracts
- **OPEN TASK**
  - We welcome any submission that uses our data!



# Impact des données partagées sur le nombre de citations



# Citations immédiates

- Automatic Simplification of Scientific Texts: SimpleText Lab at CLEF-2022** 8 2022  
L Ermakova, P Bellot, J Kamps, D Nurbakova, I Ovchinnikova, E SanJuan, ...  
European Conference on Information Retrieval, 364-373
- CLEF Workshop JOKER: Automatic Wordplay and Humour Translation** 4 2022  
L Ermakova, T Miller, O Puchalski, F Regattin, É Mathurin, S Araújo, ...  
European Conference on Information Retrieval, 355-363
- Overview of the CLEF 2022 SimpleText Lab: Automatic simplification of scientific texts** 4 2022  
L Ermakova, E SanJuan, J Kamps, S Huet, I Ovchinnikova, D Nurbakova, ...  
International Conference of the Cross-Language Evaluation Forum for European ...
- Overview of JOKER@ CLEF 2022: Automatic Wordplay and Humour Translation workshop** 9 2022  
L Ermakova, T Miller, F Regattin, AG Bosser, É Mathurin, GL Corre, ...  
Experimental IR Meets Multilinguality, Multimodality, and Interaction ...

**Merci pour votre attention !**

# Place de la donnée en sciences

Professeure Claire Hellio

Laboratoire LEMAR

IUEM

Université de Bretagne Occidentale

[claire.hellio@univ-brest.fr](mailto:claire.hellio@univ-brest.fr)

# Les scientifiques produisent des données:

- Utiles aux scientifiques
  - Innovations de ruptures
  - Avancées des connaissances
- Utiles à la Société
  - Innovations de ruptures
  - Politiques publiques
  - Avancées des connaissances
  - Sciences participatives
  - Accès aux connaissances
- Utiles au monde socio-économiques
  - Entreprises innovantes
  - Compétitivité

# Enjeux cruciaux face aux données ouvertes

- Coûts élevés de publications dans des journaux scientifiques « open access »
- Propriété intellectuelle
  - Enjeux majeur notamment dans la recherche appliquée
  - Publications vs brevets
- Assurance que les données
  - soient bien comprises par les parties prenantes
  - n'aient pas un usage détourné
  - ne soient pas plagiées

# Opportunité des données ouvertes

- Plan de gestion des données réalisés en amont des projets
- Datafication
  - Simple mesure ou transformation du réel en data qui peuvent être recueillies et analysées (rendu possible par le développement de l'informatique)
- Une donnée
  - est une information codée, figée et transmissible.
    - nécessite une codification permettant à la fois sa collecte et son rapprochement avec d'autres données formulées dans le même référentiel.
    - peut être mémorisée, pour traitement immédiat ou ultérieur.
    - Importance de l'anticipation des formats d'archivage et de l'organisation des données
- Cahier de laboratoires électroniques

# Opportunité des données ouvertes (suite)

- Accélérer l'effort de publication
- Reconnaissances des « data paper »
  - Les Digital Object Identifiers (DOI) peuvent être attribués aussi à des données/logiciels/matériels. Ainsi, si les données sont utilisées par d'autres scientifiques, les auteurs sont remerciés/cités.
  - Les articles scientifiques partageant leur matériel de recherche sont aussi plus susceptibles d'être cités, contribuant au facteur d'impact et au H Index
- Accès à des données historiques
  - Ce partage favorise les répliquions d'études
  - L'Open Data et l'Open Source sont des pratiques permettant la création de connaissances scientifiques plus robustes.
- Pérennisation des données

# Le libre accès (open access) profite directement

- aux chercheurs des pays à revenu faible ou moyen, ainsi qu'à tous ceux qui voudraient lire l'article.
- directement aux auteurs de l'article eux-mêmes car
  - leur article est directement accessible à tous dans n'importe quelle situation
    - beaucoup plus susceptible d'être lu et donc d'être cité
    - plus de citations que les articles payants
    - plus souvent partagés et discutés sur les réseaux sociaux
- **Aux entreprises**
  - Si les grandes entreprises peuvent se permettre d'avoir accès à des articles payants, il est en revanche moins probable que les petites entreprises le soient.
    - De nouveaux résultats, méthodes ou logiciels issus d'instituts de recherche ont donc moins de chances d'être adoptés par les entreprises.
    - De même, pour les chercheurs qui souhaitent promouvoir leurs travaux dans l'industrie, le fait que leur recherche soit accessible en ligne est également susceptible de promouvoir leur candidature auprès des dites entreprises.
- **Aux citoyens**
  - Etudiants, etc
  - Accès élargi aux connaissances

# Et pour aller plus loin : l'Open Review – la nouvelle tendance éditoriale

L'open Review fait référence au partage, et éventuellement à la désanonymisation, des revues (reviews) d'articles scientifiques.

- Bénéfice pour les auteurs :
  - Les revues ouvertes apportent souvent des critiques plus constructives avec des affirmations plus fondées.
  - Les rapports des référés ont tendance à être rédigés avec plus de soin.
- Bénéfice pour les référés :
  - Les revues ouvertes aident à mettre en évidence la contribution des chercheurs à l'effort d'examen par les pairs.

**Merci pour votre attention !**

# A venir

---

# A venir

- **Workshops**

- 1 par ambassadeur
- Aborder les aspects techniques et précis de l'Open Science selon les domaines de recherche
- Les dates seront communiquées sur la newsletter de la DRIVE

## Sites et liens utiles

- **CNIL** : <https://www.cnil.fr/>
- **Outil décisionnel de diffusion des données** :  
<https://sondage.univ-brest.fr/limesurvey/index.php/survey/index/sid/887695/newtest/Y/lang/fr>

# Contacts

## Les ambassadeurs

- Santé – Sofian Berrouiguet [sofian.berrouiguet@chu-brest.fr](mailto:sofian.berrouiguet@chu-brest.fr)
- SHS – Liana Ermakova [liana.ermakova@univ-brest.fr](mailto:liana.ermakova@univ-brest.fr)
- Informatique, Math appliquées, Matériaux, Radiofréquences-Télécom – Roland Gautier [roland.gautier@univ-brest.fr](mailto:roland.gautier@univ-brest.fr)
- Sciences – Claire Hellio [claire.hellio@univ-brest.fr](mailto:claire.hellio@univ-brest.fr)

## Référentes Science Ouverte à l'UBO

- Gemma Davies (DRIVE) [gemma.davies@univ-brest.fr](mailto:gemma.davies@univ-brest.fr)
- Lisa Malguy (SCD) [lisa.malguy@univ-brest.fr](mailto:lisa.malguy@univ-brest.fr)

Data Protection Officer : [dpo@univ-brest.fr](mailto:dpo@univ-brest.fr)



L-Università  
ta' Malta



## Pour en savoir plus :

- <https://sea-eu.org/researcheu/>
  - [www.univ-brest.fr/sea-eu/](http://www.univ-brest.fr/sea-eu/)
  - [www.uca.es](http://www.uca.es)
  - [www.uni-kiel.de](http://www.uni-kiel.de)
  - [www.unist.hr](http://www.unist.hr)
  - [www.um.edu.mt](http://www.um.edu.mt)
  - <https://en.ug.edu.pl/>

Contacts:

[Olivia.lahens@univ-brest.fr](mailto:Olivia.lahens@univ-brest.fr)



Co-funded by the  
Erasmus+ Programme  
of the European Union



**« Aussi ouvert que possible, aussi fermé que nécessaire »**