

# Usages *alpha*

## *Étude préliminaire à orientation méthodologique*

**Porteur de projet** : Marin Dacos, directeur du Centre pour l'édition électronique ouverte

**Partenaires du projet** :

- Joël Gombin, SCOP Dataactivist
- Pierre-Carl Langlais, Post-doctorant SIC

Cette recherche a reçu le soutien du Programme d'investissement ISTEEX  
(ANR-10-IDEX-0004-02)

**ISTEX**

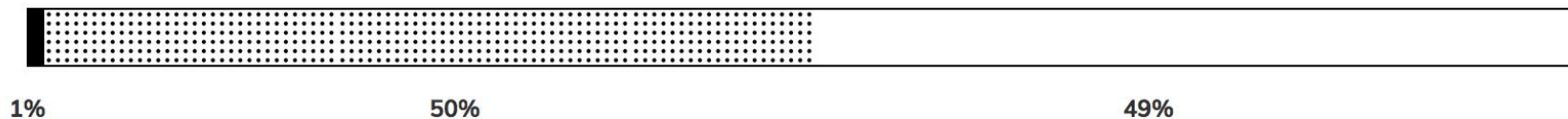
Journées ISTEEX

15-16 mars 2018 Paris (France)



# Contexte du projet

How many citations are open today?

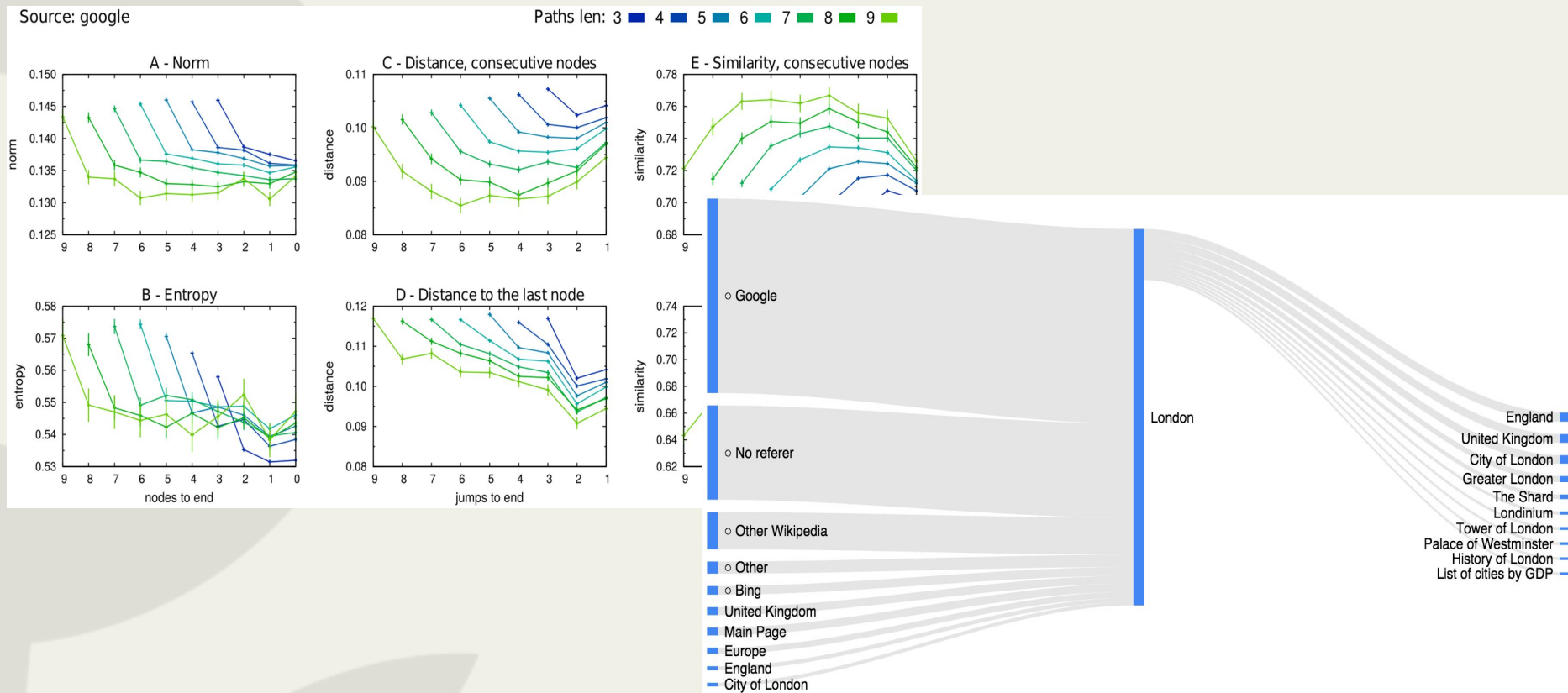


As of January 2018, the fraction of publications with open references has grown from 1% to more than 50% out of 38 million articles with references deposited with Crossref.

We encourage all other scholarly publishers to follow the example of these trail-blazing publishers by making their reference metadata publicly available. Please contact Crossref Support ([support@crossref.org](mailto:support@crossref.org)) for more information, or to let them know that you are ready to open up your reference metadata now. See also our list of responses to [frequently asked questions](#).

Depuis tout juste un an les mesures de la réception scientifique s'ouvrent massivement : il y a aujourd'hui 50% de citations ouvertes sur Crossref.

# Contexte du projet

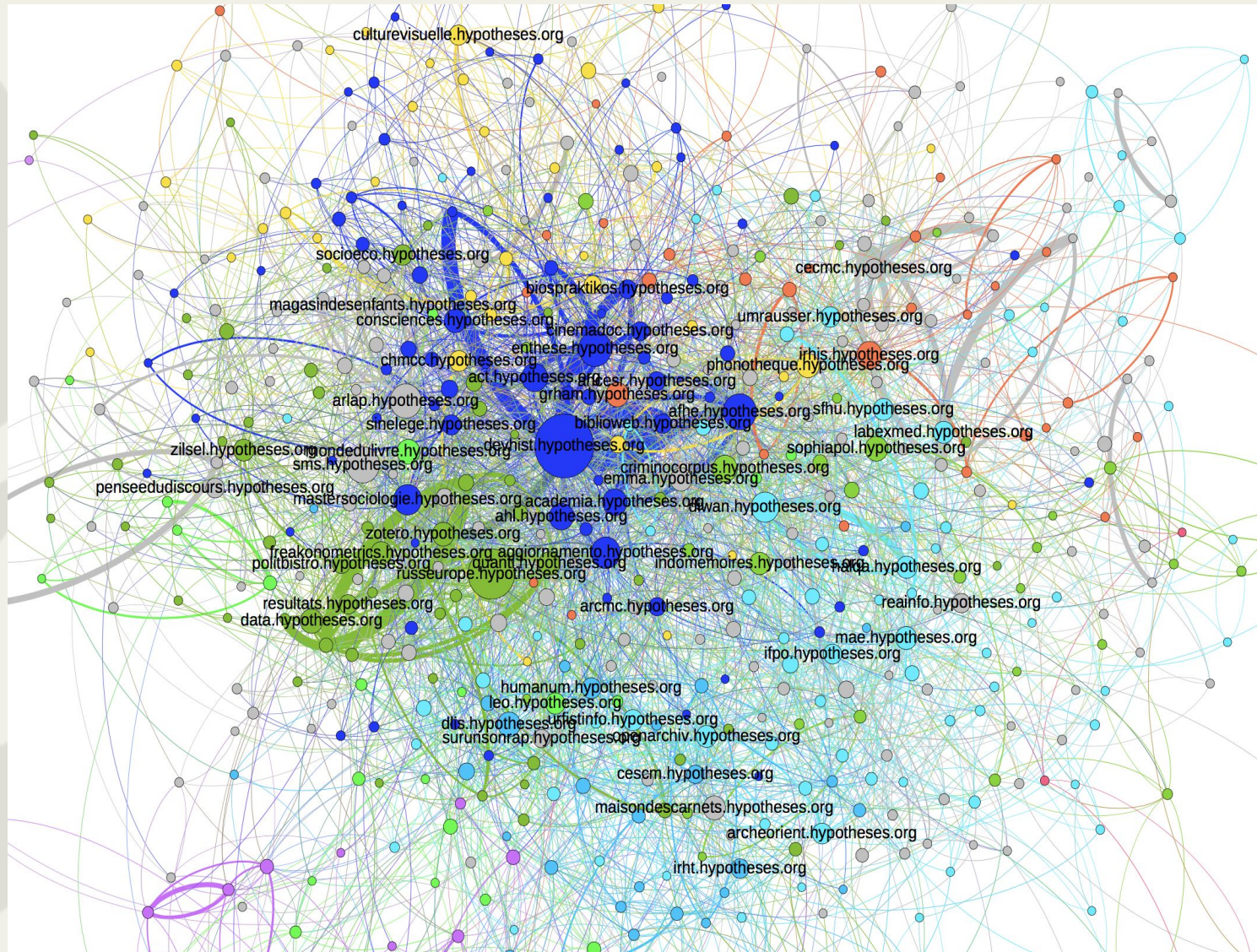


La citation suffit-elle à représenter toutes les formes de réception de publications scientifiques en libre accès ? Et si l'on s'appuyait également sur les parcours des lecteurs.

# Contexte du projet



# Contexte du projet



# Présentation du projet

## **Objectif :**

Définir une grille d'analyse par un dispositif méthodologique pour identifier les usages de lecture sur les plateformes d'OpenEdition

## **Deux principales réalisations :**

- le développement d'un détecteur des usages présents sur la plateforme OpenEdition sous la forme d'une application Web,
- définition d'une grille d'analyse des usages des lecteurs sur les plateformes d'OpenEdition.

# Corpus et méthodologies

Type de statistiques	URL	Descriptif
Statistiques d'accès	<a href="https://logs.openedition.org">https://logs.openedition.org</a>	Mesure de la fréquentation des sites : visiteurs, pages vues, trafic
Statistiques d'usage(s)	<a href="http://usages.openedition.org">http://usages.openedition.org</a> (Accès intranet)	Statistiques de consultation des articles ou chapitres par numéros ou livres, par auteurs, par origines (referers ou affluents) sur une période paramétrable
Piwik		Logs de connexion (60Go Mysql pour 12 mois)

Corpus	Volume
Revue.org	190.000 documents
Hypothèses	230.000 billets
Calenda	34.000 annonces
OpenEdition Books	6.500 documents

# Corpus et méthodologies

## Un projet de *Little Big Data*

- **Exploration de l'ensemble des logs** de janvier à mi-novembre 2017, soit **77 millions de visites** réparties en **47 millions de sessions**. C'est un élargissement significatif du projet initial qui prévoyait des échantillons sur une durée de quelques jours.
- **Mise en visibilité de nouveaux objets** : l'utilisation des logs complet rend possible le repérage du lecteur inattendu et la réalisation de segmentations fines.
- **Contraintes logistiques** : sans être considérable, le volume des données étudiées **met à l'épreuve les outils d'analyse usuels** et elles sont en évolution constante. les temps de traitement sont substantiellement allongés.
- **Données complexes à modéliser** : des modèles inhabituels sont nécessaires pour rendre compte de la diversité des phénomènes.



# Sommaire

1. Détecter les lecteurs inattendus
2. Cartographier les parcours



# 1. Détecter les lecteurs inattendus

*De l'identification de cas particulier à la  
découverte d'un phénomène global*

# 1. Lecteurs inattendus

## L'émergence du lecteur inattendu

OPEN ACCESS AND INSTITUTIONAL REPOSITORIES, SCHOLARLY PUBLISHING

### The unexpected reader

🕒 NOVEMBER 15, 2011 👤 KEVIN SMITH, J.D. 💬 4 COMMENTS

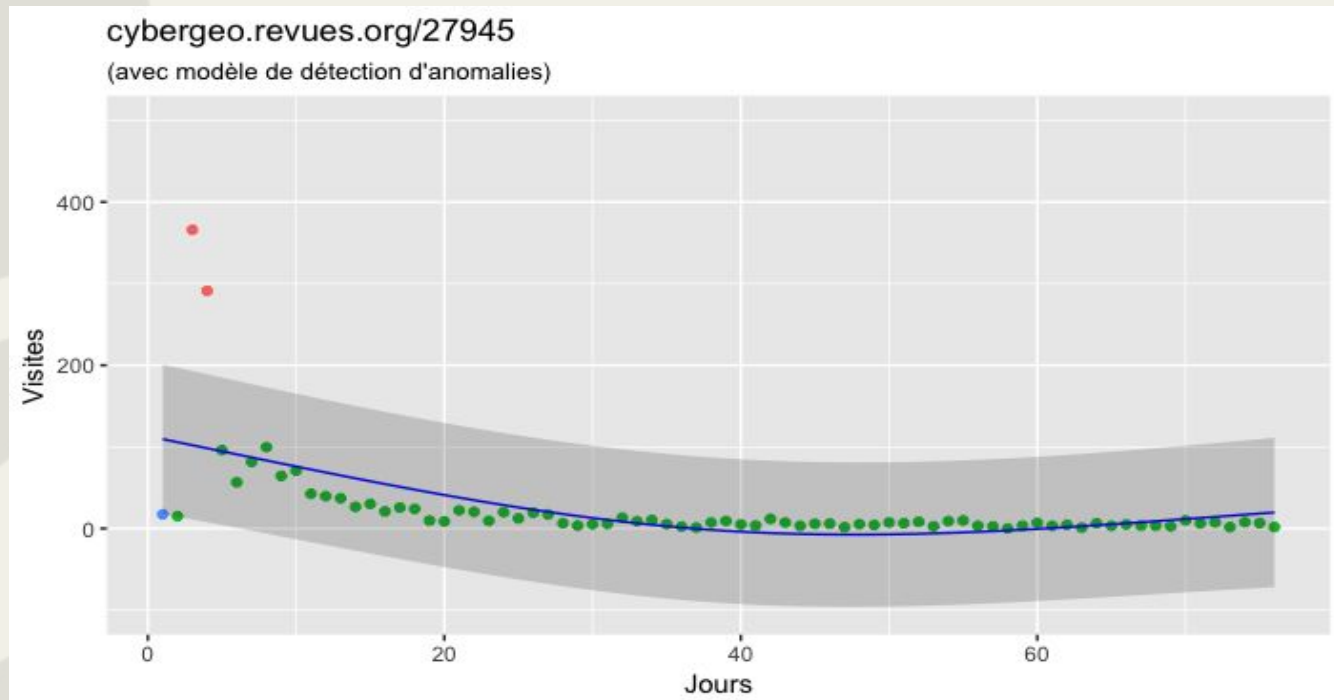
I have just returned from the Berlin 9 Conference on Open Access, which was held in Washington, D.C. at the lovely conference center facilities of the Howard Hughes Medical Institute. It was a fascinating meeting, and quite different in tone from the one I attended last year in Beijing.

In its opening paragraph, this [Chronicle of Higher Education report](#) on the conference captures the fundamental difference. This year the conference was much more clearly focused on the *impact* of open access on research; rather than talking about how open access will be accomplished, the discussion assumed that open access is inevitable and instead emphasized the differences that the evolution to open will make.

Une nouvelle forme de lecture rendue possible par le libre accès : la diffusion n'est plus uniquement restreinte aux pairs

# 1. Lecteurs inattendus

## Modéliser le lecteur inattendu



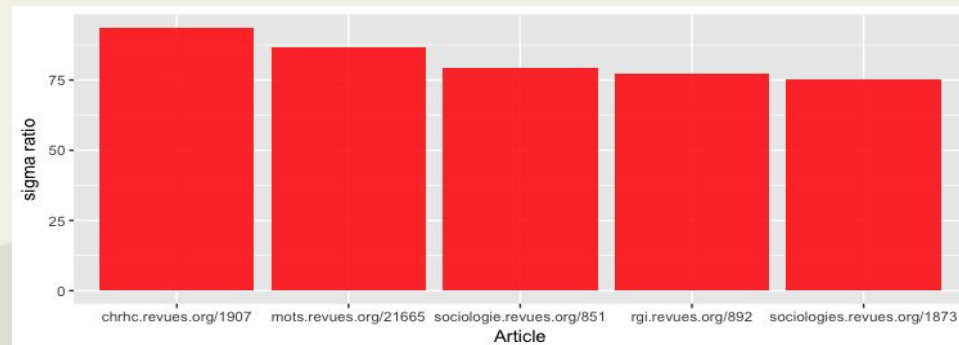
Le modèle représente les évolutions *structurelles* de la fréquentation

Ce qui en déborde est “inattendu”

# 1. Lecteurs inattendus

## Modéliser le lecteur inattendu

```
> anomaly_detection@url_prediction_anomalies %>% arrange(desc(fit_sigma_ratio)) %>% select(-lwr, -upr)
# A tibble: 32,087 × 5
  url diff_days log_value_anoms fit fit_sigma_ratio
  <chr> <dbl> <dbl> <dbl> <dbl>
1 chrhc.revues.org/1907 82 1748.3 40.590581 93.69053
2 mots.revues.org/21665 50 197.9 8.253494 86.53669
3 sociologie.revues.org/851 24 144.5 7.647771 79.31815
4 rgi.revues.org/892 97 168.5 8.926589 77.46158
5 sociologies.revues.org/1873 34 265.8 13.085899 75.29770
6 confins.revues.org/11729 6 2615.8 272.741627 74.65389
7 insaniyat.revues.org/5292 83 226.0 11.101971 73.66207
8 confins.revues.org/11701 10 183.5 15.243524 73.32456
9 com.revues.org/5838 36 294.4 19.330700 72.79988
10 terrain.revues.org/4263 31 98.0 6.078802 72.54528
```



Notre modèle permet de repérer les anomalies les plus extrêmes sur l'ensemble de revue.org

# 1. Lecteurs inattendus

## *Commémorer*

### Une commémoration

Hassan Fathy, construire *avec* ou *pour* le peuple?

Thierry Paquot

p. 15-25

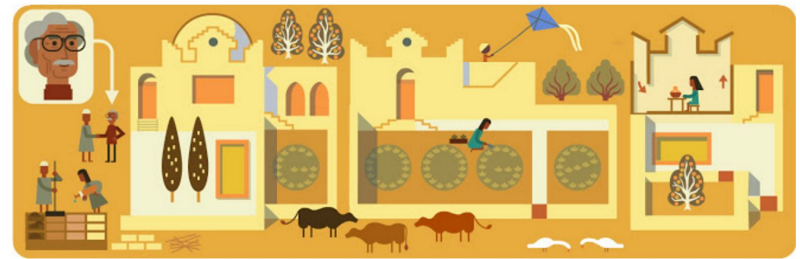


Résumé | Index | Plan | Texte | Notes | Citation | Auteur

#### Résumé

Au lendemain de la Seconde Guerre mondiale, l'architecte égyptien Hassan Fathy, en se lançant dans l'aventure architecturale du village de Gournah, propose des constructions économes et expérimente de nouveaux procédés constructifs. Militant de l'auto-construction et d'une esthétique de qualité, il associe les peuples à la construction de leurs maisons et exploite les traditions locales. La réception de son œuvre pose la question d'une architecture pour le peuple et d'une rupture douce avec le passé.

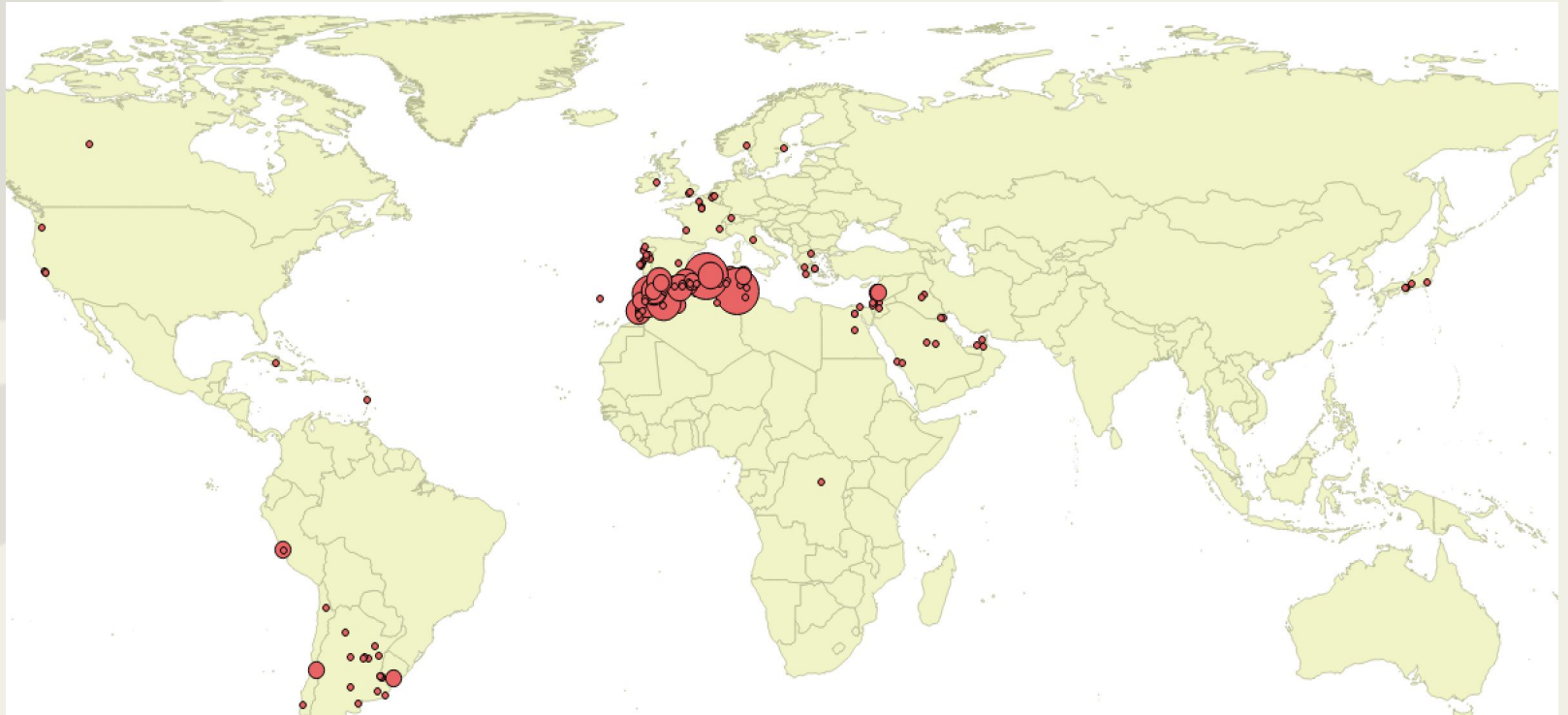
<http://chrhc.revues.org/1907>



2000 visites en une journée pour un vieil article qui n'en a que 3 ou 4 habituellement. La cause : un doodle de Google consacré à l'architecte égyptien Hassan Fathy

# 1. Lecteurs inattendus

## *Commémorer*

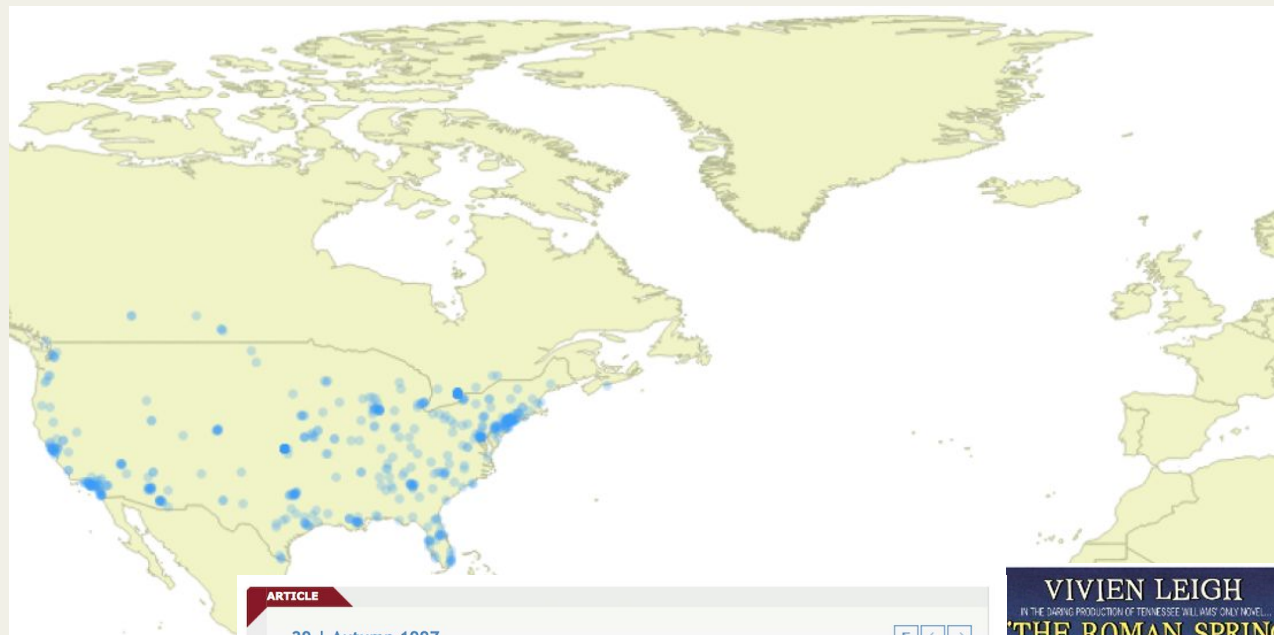


Les coordonnées des visites suggèrent que le doodle a surtout été activé dans les pays du Maghreb

# 1. Lecteurs inattendus

## *Prolonger un film...*

Hassan Fathy est un cas exceptionnel mais il y a des anomalies beaucoup plus fréquentes : les articles sur des films ont plein de visites... après leur diffusion à la télévision.



ARTICLE

29 | Autumn 1997 ⏪ ⏩ ⏴ ⏵

Other places, other selves

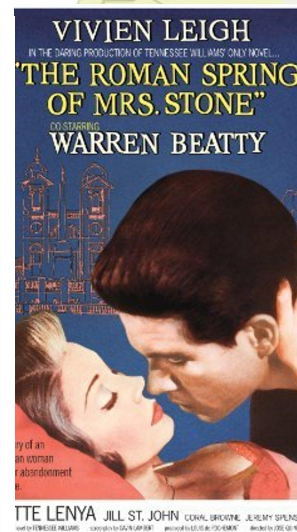
**The kindness of strangers : savage rescue in Tennessee**  
**Williams' *The roman spring of mrs. Stone***

Theoharis Theoharis Constantine

Résumé | Texte | Citation | Auteur

**RÉSUMÉ**

Dans la longue nouvelle "Le Printemps romain de Mrs Stone", Williams présente l'éveil tardif à la vie d'une actrice américaine qui vit loin de chez elle. A travers des stratégies narratives qui engagent l'ambiguïté, des modes de construction de personnages fondés sur le paradoxe, l'auteur fait traverser à son héroïne deux amours sacrificielles : en passant d'un ami entretenu attiré à un jeune garçon qui se prostitue dans la rue, Mrs Stone laisse derrière elle les conventions d'une vie élégamment décadente pour pénétrer dans le violent univers du paganisme que le titre de la nouvelle semble suggérer. Sa délivrance est, certes, crue, teintée de barbare. Cette lecture psychanalytique de l'identité sexuelle du personnage est fondée sur la conception classique de la représentation du désir.

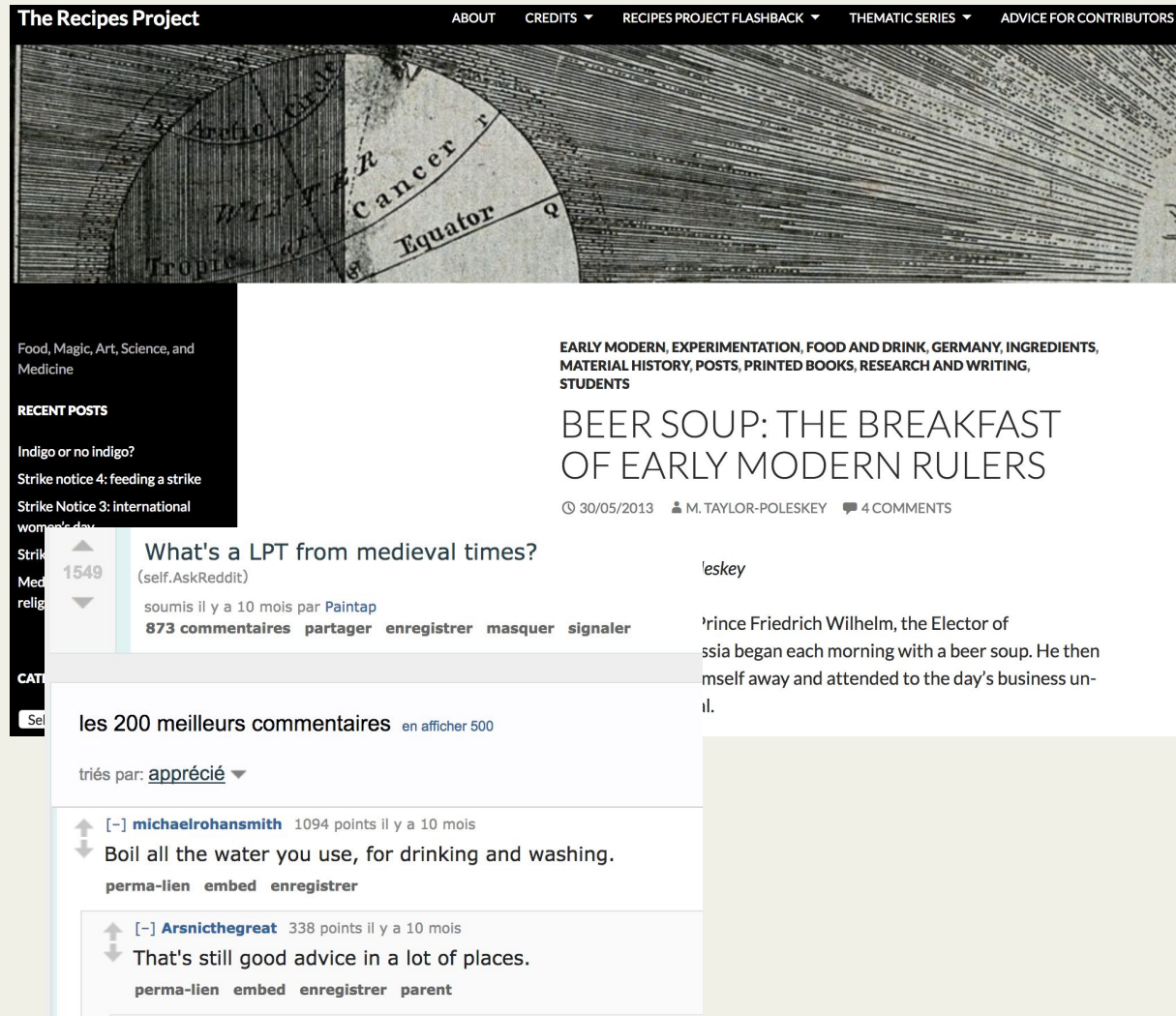




# 1. Lecteurs inattendus

## *De la science à l'anecdote à la science*

De nombreux épisodes de fréquentation inattendu mettent en évidence le rôle de nouveaux intermédiaires entre publications scientifiques et lecteurs inattendus



The Recipes Project

ABOUT CREDITS RECIPES PROJECT FLASHBACK THEMATIC SERIES ADVICE FOR CONTRIBUTORS

Food, Magic, Art, Science, and Medicine

RECENT POSTS

Indigo or no indigo?

Strike notice 4: feeding a strike

Strike Notice 3: international women's day

Strike

Med

relig

CAT

Sel

EARLY MODERN, EXPERIMENTATION, FOOD AND DRINK, GERMANY, INGREDIENTS, MATERIAL HISTORY, POSTS, PRINTED BOOKS, RESEARCH AND WRITING, STUDENTS

### BEER SOUP: THE BREAKFAST OF EARLY MODERN RULERS

30/05/2013 M. TAYLOR-POLESKEY 4 COMMENTS

What's a LPT from medieval times?  
(self.AskReddit)  
soumis il y a 10 mois par Paintap  
873 commentaires partager enregistrer masquer signaler

leskey

Prince Friedrich Wilhelm, the Elector of Prussia began each morning with a beer soup. He then fell asleep and attended to the day's business until.

les 200 meilleurs commentaires en afficher 500

triés par: apprécié

[-] michaelrohansmith 1094 points il y a 10 mois  
Boil all the water you use, for drinking and washing.  
perma-lien embed enregistrer

[-] Arsnicthegreat 338 points il y a 10 mois  
That's still good advice in a lot of places.  
perma-lien embed enregistrer parent

# 1. Lecteurs inattendus

## De la science à l'anecdote à la science

De nombreux épisodes de fréquentation inattendu mettent en évidence le rôle de nouveaux intermédiaires entre publications scientifiques et lecteurs inattendus

Accueil › High-Tech / Science › Top 10 des preuves historiques qu'on vit dans un monde parallèle, la SF dans la vraie vie

[TOP PRÉCÉDENT](#)  
10 explications possibles derrière ...

[TOP SUIVANT](#)  
10 villes des Etats-Unis à voir a ...

### Top 10 des preuves historiques qu'on vit dans un monde parallèle, la SF dans la vraie vie

Par **Pierre Galouise**  
le 4/04/2017

**699**  
PARTAGES

Catégorie : [High-Tech / Science](#)

Entre quelques **phénomènes météorologiques** au fil de l'Histoire, on a parfois l'impression que le top a principalement été inspiré par l'**la mesure du temps** » qui nous livre et c'est DINGO.

**Index**

- Auteurs
- Mots-clés
- Chronologique
- Thématique

**Derniers numéros**

- 24-25 | 2002

**Numéros en texte intégral**

- 45 | 2014
- 44 | 2013
- 42-43 | 2012
- 41 | 2010
- 40 | 2009
- 39 | 2008
- 38 | 2008
- 36-37 | 2007

### Les chemins de fer et l'heure légale

**Lucien Baillaud**

Plan | Texte | Notes | Citation | Auteur

#### Plan

- Les inconvénients des heures locales
- Vers l'unification des heures : les heures nationales
- L'heure de Rouen
- Le système américain
- La conférence de Washington
- Vers la généralisation des fuseaux horaires
- 24 heures ? ou 12 et 12 ?
- La décimalisation du temps et les problèmes de minutes
- Conclusion



# 1. Lecteurs inattendus

Au-delà des cas particuliers ici mis en lumière, la lecture inattendue est une pratique courante pour les publications en libre accès. Au-delà de leur réception par les pairs, les articles peuvent :

- Préserver et documenter une mémoire collective
- Prolonger la découverte d'une production culturelle (film, livre, etc.)
- Recontextualiser la vulgarisation scientifique.

# 1. Lecteurs inattendus

## *L'application Umberto*

Umberto présente deux écrans successifs.

- le premier écran permet de sélectionner l'épisode de lecteurs inattendus que l'on souhaite explorer.
- le second écran présente les informations disponibles sur l'épisode de lecteur inattendu retenu.

# 1. Lecteurs inattendus

## *L'application Umberto*

Détecteur de lecteurs inattendus

https://rstudio.labocleo.org/p/4122/

**OpenEdition**

Bienvenue dans le détecteur de lecteurs inattendus !

Plateformes :

- ON Hypothèses
- ON Revues.org
- ON OpenEdition Books
- ON Calenda

Publications :

Choisir un ou plusieurs espaces de publications :

Période :

Choisir la période concernée

01/01/2017 à 13/11/2017

Intensité minimale

0 0.2 1

Choisissez l'épisode de lecteurs inattendus que vous voulez explorer

Afficher 10 éléments

Rechercher :

	Plateforme	Publication	Document	URL	Intensité	Date
1	Revue.org	Entrelacs	Matthew Barney, artiste/auteur : pour une hybridation des genres1	http://entrelacs.revues.org/343	97	04/10/2017
2	Revue.org	Études de communication	Technologies de l'information et de la communication et modernisation des services publics. Quelques remarques et repères pour une exploration critique	http://edc.revues.org/1135	95	25/09/2017
3	Revue.org	Genre, sexualité et société	Virginité des filles et rapports sociaux de sexe dans quelques récits d'écrivaines marocaines contemporaines	http://gss.revues.org/1413	93	22/04/2017
4	Revue.org	Cahiers d'histoire. Revue d'histoire critique	Hassan Fathy, construire avec ou pour le peuple?	http://chrhc.revues.org/1907	92	24/03/2017
5	Revue.org	Sociologie	Communauté et société	http://sociologie.revues.org/2574	91	23/07/2017
6	Hypotheses.org	Archivalia	Open Access Woche (Tag 2)	http://archivalia.hypotheses.org/68021	91	25/10/2017
7	Hypotheses.org	Zilsel	Réponse à la réponse de Pascal Ragouet	http://zilsel.hypotheses.org/2928	90	23/10/2017
8	Revue.org	Insaniyat / إنسانيات	2002-1963) من الثورات الزراعية إلى الإصلاحات الليبرالية (الفاحة في الجزائر)	http://insaniyat.revues.org/7027	87	30/06/2017
9	Revue.org	Cahiers de la Méditerranée	Les Alpes-Maritimes et la crise franco-monégasque de 1962	http://cdlm.revues.org/3023	87	22/08/2017
10	Hypotheses.org	Le blog du collectif RENOIR	Offre de stage "Sauvegarde du patrimoine de l'observatoire du Pic du Midi"	http://renoir.hypotheses.org/1946	86	07/10/2017

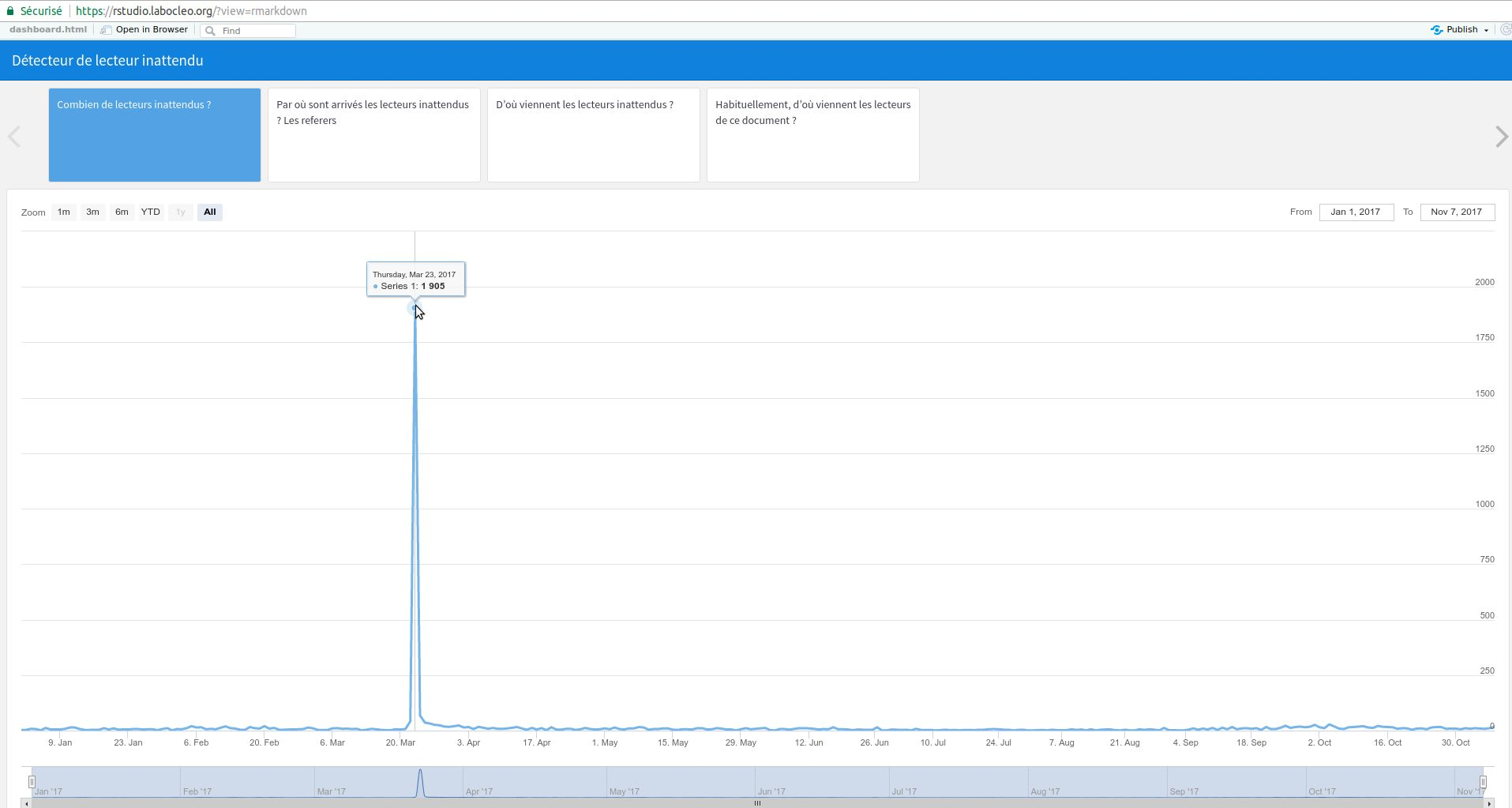
Affichage de l'élément 1 à 10 sur 479 éléments

Précédent 1 2 3 4 5 ... 48 Suivant

Explorez !

# 1. Lecteurs inattendus

## *L'application Umberto*



# 1. Lecteurs inattendus

## *L'application Umberto*

Sécurisé | <https://rstudio.labocleo.org/?view=rmarkdown>

dashboard.html | Open in Browser | Find

Publish

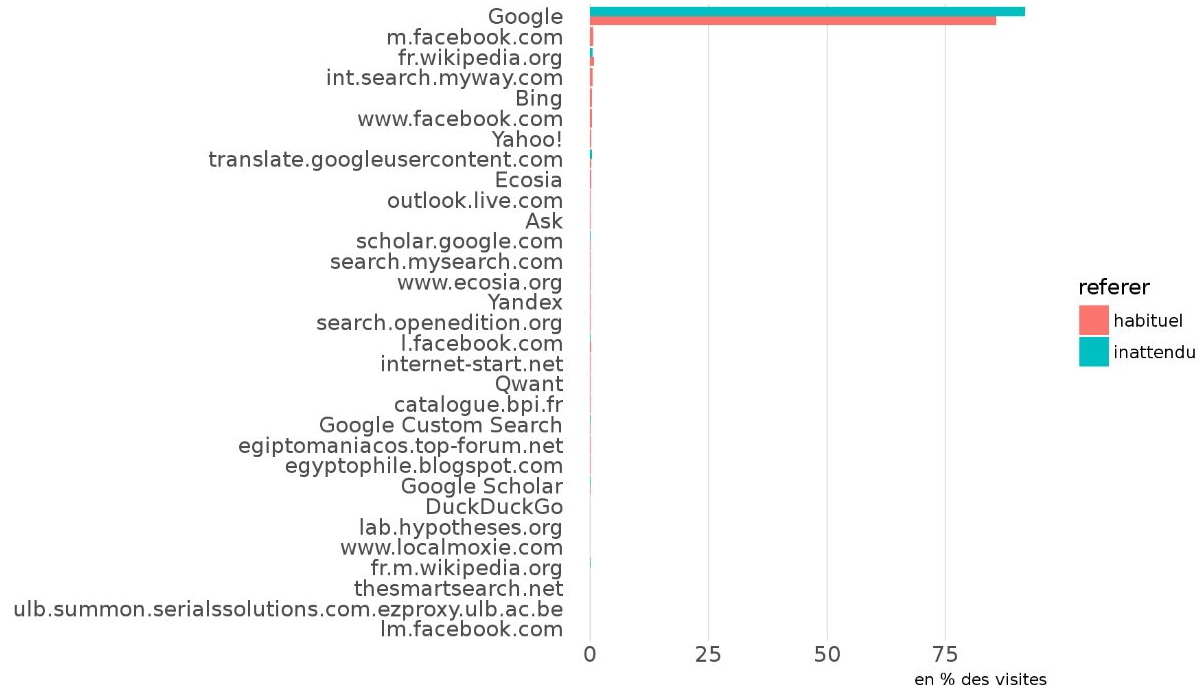
Détecteur de lecteur inattendu

Combien de lecteurs inattendus ?

Par où ont arrivés les lecteurs inattendus ?  
? Les referers

D'où viennent les lecteurs inattendus ?

Habituellement, d'où viennent les lecteurs de ce document ?





# 1. Lecteurs inattendus

## *L'application Umberto*

Sécurisé | <https://rstudio.labocleo.org/?view=rmarkdown>

dashboard.html | Open in Browser | Find

Publish

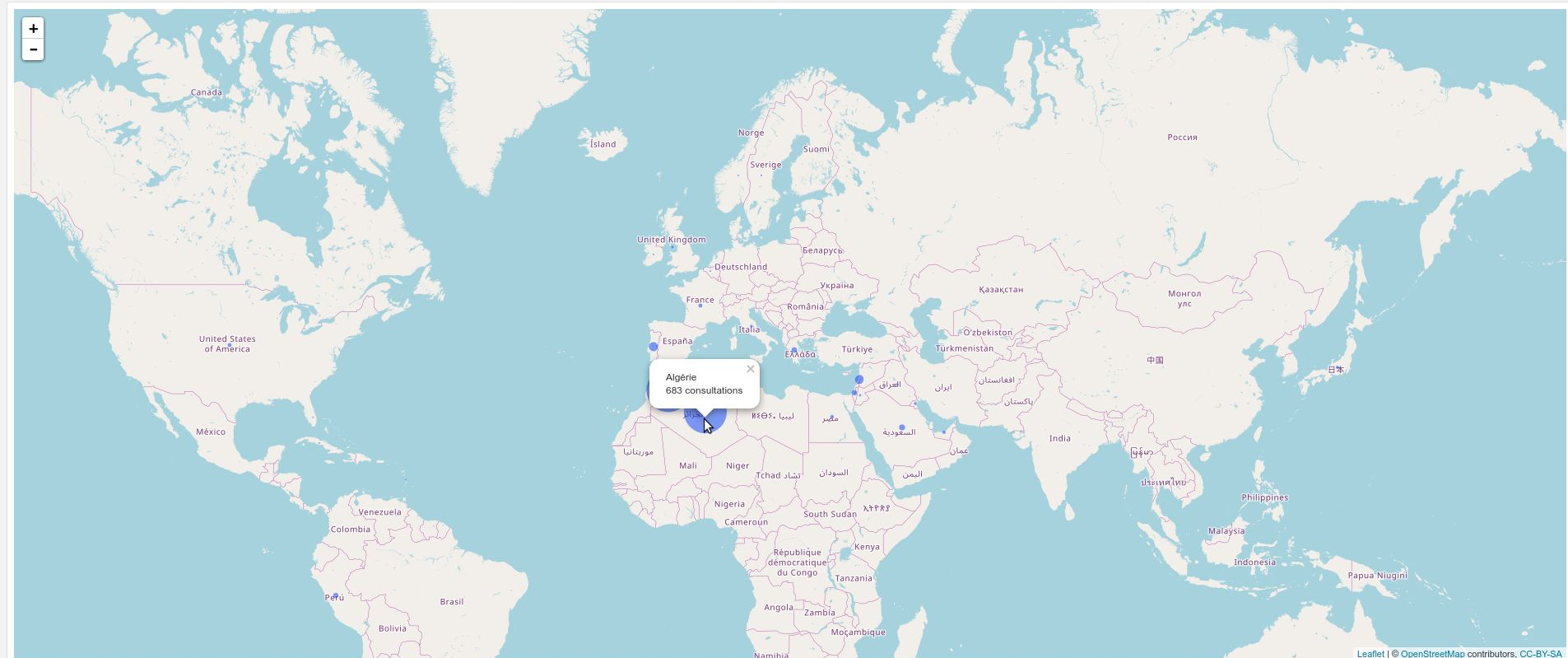
Détecteur de lecteur inattendu

Combien de lecteurs inattendus ?

Par où sont arrivés les lecteurs inattendus ?  
Les referers

D'où viennent les lecteurs inattendus ?

Habituellement, d'où viennent les lecteurs de ce document ?





## 2. Cartographier les parcours

*Des pratiques usuelles du lecteur  
attendu.*



## 2. Cartographier les parcours

### *Des consultations en réseau*

#### Une taxinomie des formes de lecture

La classification des parcours repose sur l'identification automatisée de 6 formes de lecture appliquée à

- **La lecture « augmentée »** : ouverture d'images et, dans une moindre mesure, de fichiers intégrés à une publication.
- **La lecture locale** : ouverture de liens présents dans un article conduisant vers un autre article d'une même publication
- **La lecture bibliographique** : ouverture de liens présents dans un publication conduisant vers des publications extérieures
- **La lecture extérieure** : consultation de publications extérieures à OpenEdition qui ne sont pas liées dans la publication.
- **La lecture par sommaire/menu** : utilisation régulière de la page d'accueil de la revue / du carnet au cours de la session.
- **La lecture par recherche** : utilisation régulière des moteurs de recherche au cours de la session.

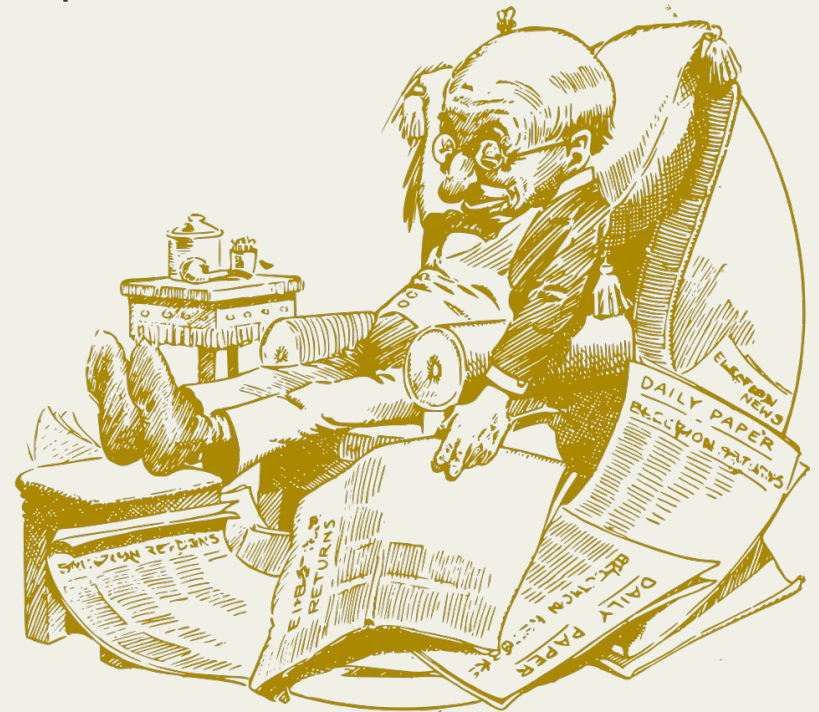
## 2. Cartographier les parcours *Des consultations en réseau*

### Une taxinomie des formes de lecture

Certaines formes de lectures (augmentée, locale, bibliographique et extérieure) sont subdivisées en deux variantes : en regard et compulsive.



*Lecture en regard*



*Lecture compulsive*

# 2. Cartographier les parcours

## *Des consultations en réseau*

### Une taxinomie des formes de lecture

Un exemple de lecture en regard : plusieurs dizaines d'aller-retour entre un carnet de recherche sur l'héraldique et un ouvrage de 1861 (!) numérisé sur Gallica sur les armoiries de la noblesse européenne.

**Heraldica**  
NOVA  
MEDIEVAL AND EARLY MODERN HERALDRY FROM THE PERSPECTIVE OF CULTURAL HISTORY

ARTICLES / RESOURCES & TOOLS

About the Blog | How to Contribute | English | Deutsch | Français | Credits | ISSN: 2199-0816 | Login for Contributors

### Databases & Materials

BY HERALDICA NOVA · PUBLISHED 01/01/2014 · UPDATED 16/06/2017

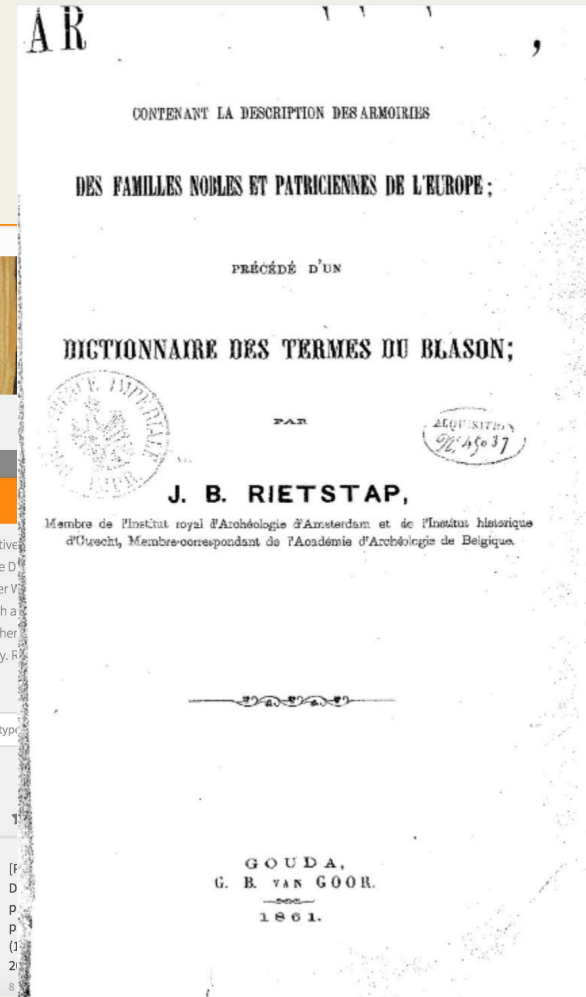
These databases and dictionaries will help to identify a depicted coat of arms as well as a blasoned one.

#### Databases

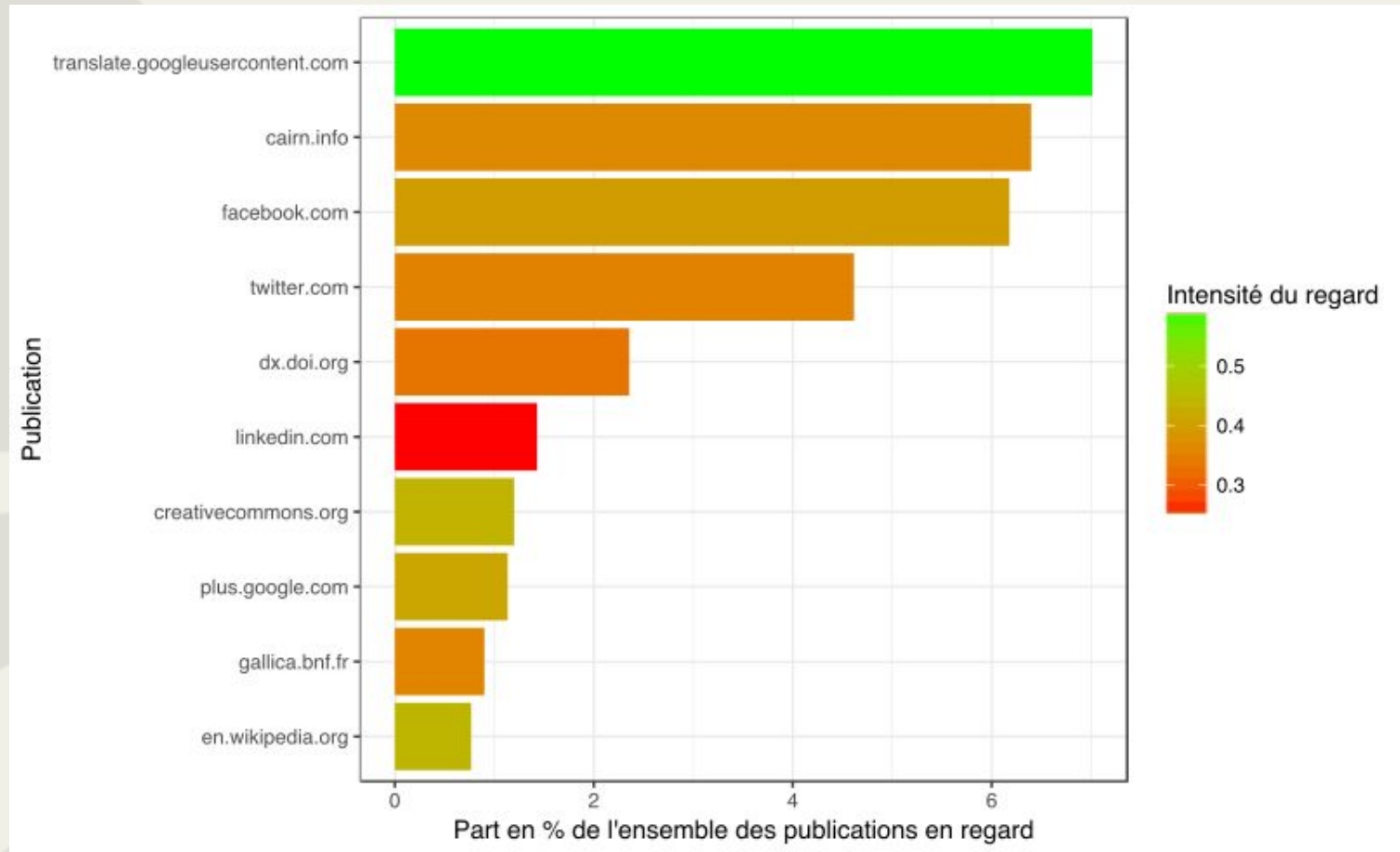
##### Steen Clemmensen, Medieval Armorial

This website is dedicated to armorials. Clemmensen has made an elaborate database of armorials and coats of arms in armorials, which can be downloaded from the website, containing more than 14,000 individual coats of arms. This database enables you, among other things, to identify coats of arms and find out which family bore which coat of arms in the Middle Ages. Furthermore, it presents quality editions of armorials by the author which can be consulted for free.

For more information on this database: <http://heraldica.hypotheses.org/673>.  
Weblink: <http://www.armorial.dk/>



## 2. Cartographier les parcours *Des consultations en réseau*



Les indicateurs de formes de lecture font émerger des pratiques insoupçonnées :  
Google Traduction est fréquemment utilisé pour une lecture en “aller-retour”

## 2. Cartographier les parcours *Des consultations en réseau*

### Une taxinomie des formes de lecture

Un cas de lecture compulsive :  
ouverture de plusieurs dizaines  
de lien présents dans un article sur  
des artistes coréens contemporains.

Art Paris Art Fair 2017 : Artistes  
coréens, jusqu'au 2 avril

PAR ARTSDELACOREE · 31/03/2017



**Art Paris Art Fair**

Grand Palais

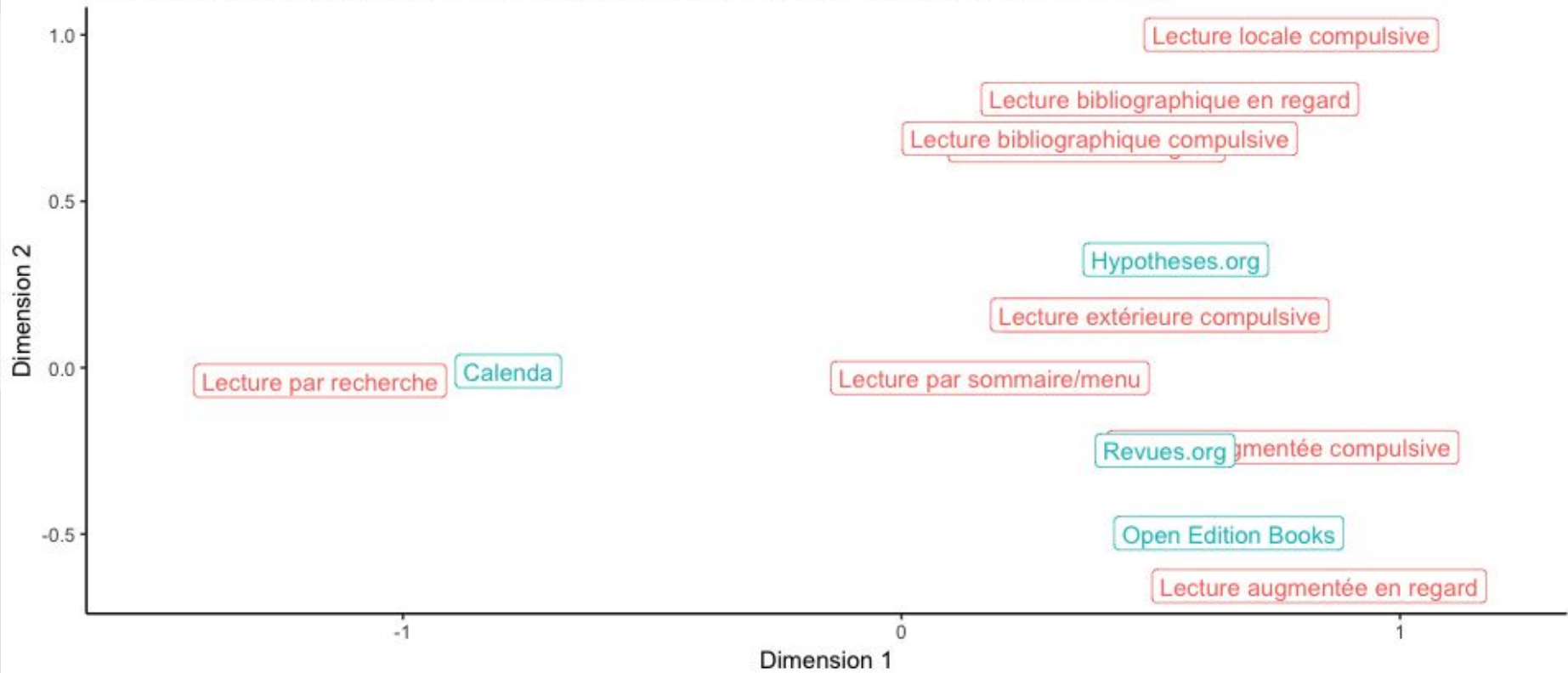


## 2. Résultats obtenus

### *Observation des usages*

#### Analyse de correspondances des parcours d'Open Edition

On lit les références d'Hypotheses, les documents additionnels de Revues et Books et on recherche sur Calenda



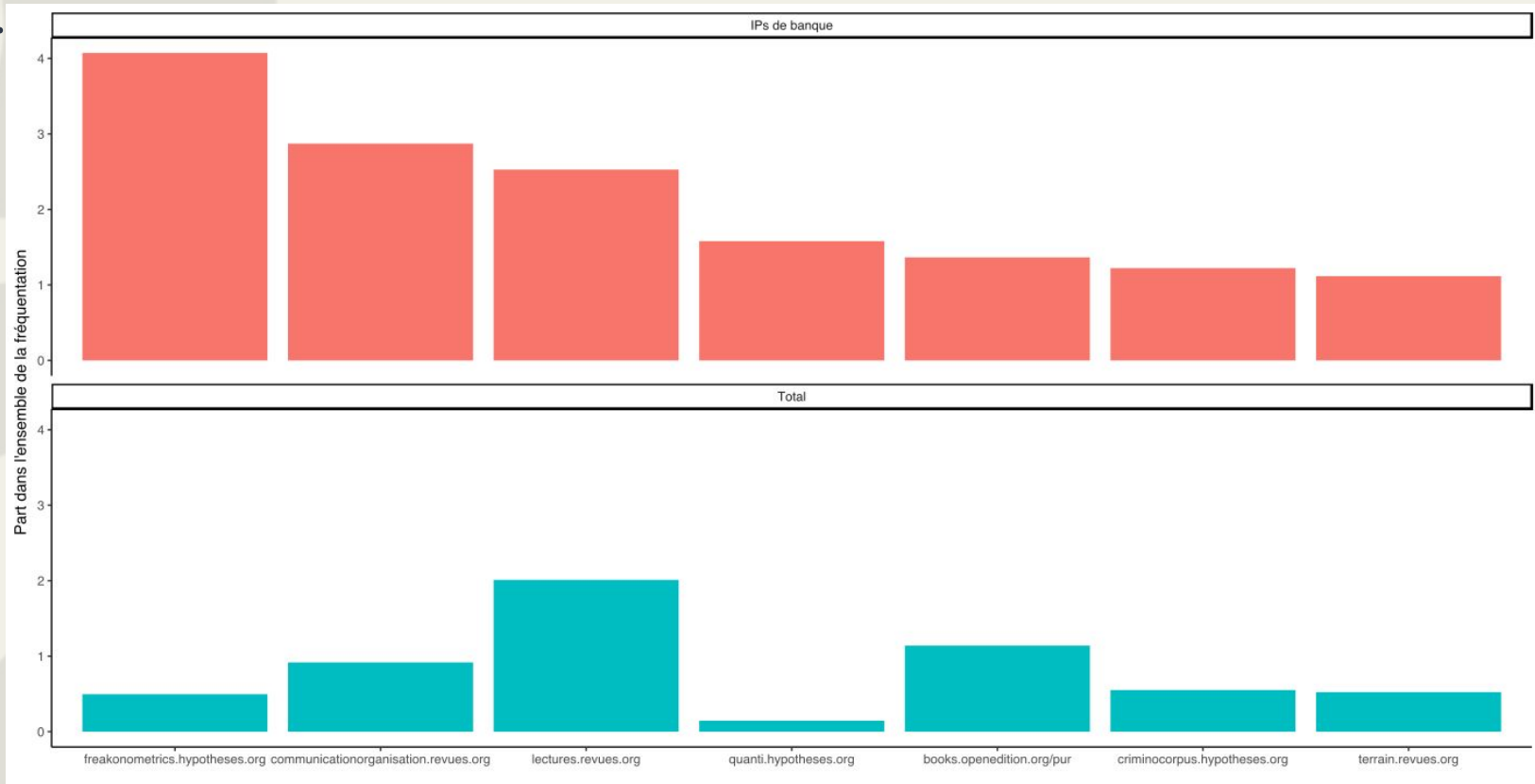
Les formes de lecture sont conditionnées par le dispositif éditorial : on lit des sources sur hypotheses ; on regarde des fichiers additionnels sur Revues ; on recherche sur Calenda.

## 2. Résultats obtenus

### *Observation des usages*

## Un lecteur institutionnel inattendu : les banques

Parmi les organisations hors recherche, les banques sont très présentes avec plusieurs milliers de sessions ouvertes par la BNP, le Crédit Mutuel, le Crédit Agricole.



## 2. Résultats obtenus

### *Observation des usages*

## Un lecteur institutionnel inattendu : les banques

Parmi les organisations hors recherche, les banques sont très présentes avec plusieurs milliers de sessions ouvertes par la BNP, le Crédit Mutuel, le Crédit Agricole.





# Conclusion

*Ce n'est que le début...*

# Conclusion

- Pour prolonger la réflexion, plusieurs pistes d'étude :
  - Repérer les similarités thématiques dans les parcours
  - Identifier les modalités de la citation à partir des termes qui accompagnent un lien
  - Affiner la typologie des parcours
- Exploitation des résultats dans le cadre du projet Appropriations financés par la fondation A\*MIDEX

# Communication

## Plusieurs communications :

Pierre-Carl Langlais, *Modéliser la lecture - Une approche quantitative de la consultation des revues scientifiques en accès ouvert*, Séminaire DH, EHESS, Juin 2017

<https://drive.google.com/file/d/0B775SZpk-FplbFlzd28xSlJjSXM/view?usp=sharing>

Marin Dacos et al., “The unexpected reader”, *DH2017 Conference keynote* – 8th August 2017

<https://dh2017.adho.org/program/plenieres/?lang=fr>

Présentation de cas emblématiques d’usage inattendu :

<http://lab.hypotheses.org/category/etude-sur-les-usages-des-lecteurs>

Pierre - Carl Langlais, Retours d’expérience : Evolution de la communication scientifique. Journée « Innométriques 2017 », CNRS.

[https://drive.google.com/file/d/1jP-tAM\\_JniE3cPkagx-15S5mQAAdY8sw/view](https://drive.google.com/file/d/1jP-tAM_JniE3cPkagx-15S5mQAAdY8sw/view)

# Communication

## Communications envisagées :

- publication sur le carnet d'OpenEdition Lab (<http://lab.hypotheses.org/>) de deux billets portant sur le prototype de détecteur de lecteurs inattendus et sur les analyses d'usages observés sur la plateforme OpenEdition
- réponse à deux appels à contribution :
  - LIBER's 2018 Annual Conference (<http://libereurope.eu/blog/2017/11/24/liber-2018-call-papers/>)
  - The 22nd International Conference on Electronic Publishing (EIPub 2018) (<http://epress.utsc.utoronto.ca/elpub2018/call-for-papers/>)
- rédaction d'un article scientifique en français et en anglais