

# Archives ouvertes et statistiques de consultations : des indications alternatives ?

Pier Luigi ROSSI (IRD)

# Le fonds documentaire de l'IRD, pionnier des archives institutionnelles et ouvertes ?

- ◊ Une archive institutionnelle créée en 1955.
- ◊ Une base bibliographique informatisée dès 1986 : base Horizon, 98.000 références documentaires. « Pleins » de champs de « métadonnées » : auteurs, date de publication, thématiques, langue du document, périodique/collection, ...
- ◊ Numérisation, texte intégral et libre accès dès 1996 : projet Pleins\_Textes . 65.000 documents au format PDF en libre accès sur internet.

# Que faire avec une base bibliographique et des « human accesses » ?

- ◆ Des tableaux de bord dynamiques en croisant les champs des métadonnées et les champs des logs

## Comment faire ?

- ◆ Filtrage en mode ligne de commande, php, mysql, librairies de graphes dynamiques (protovis, highcharts, sortable )

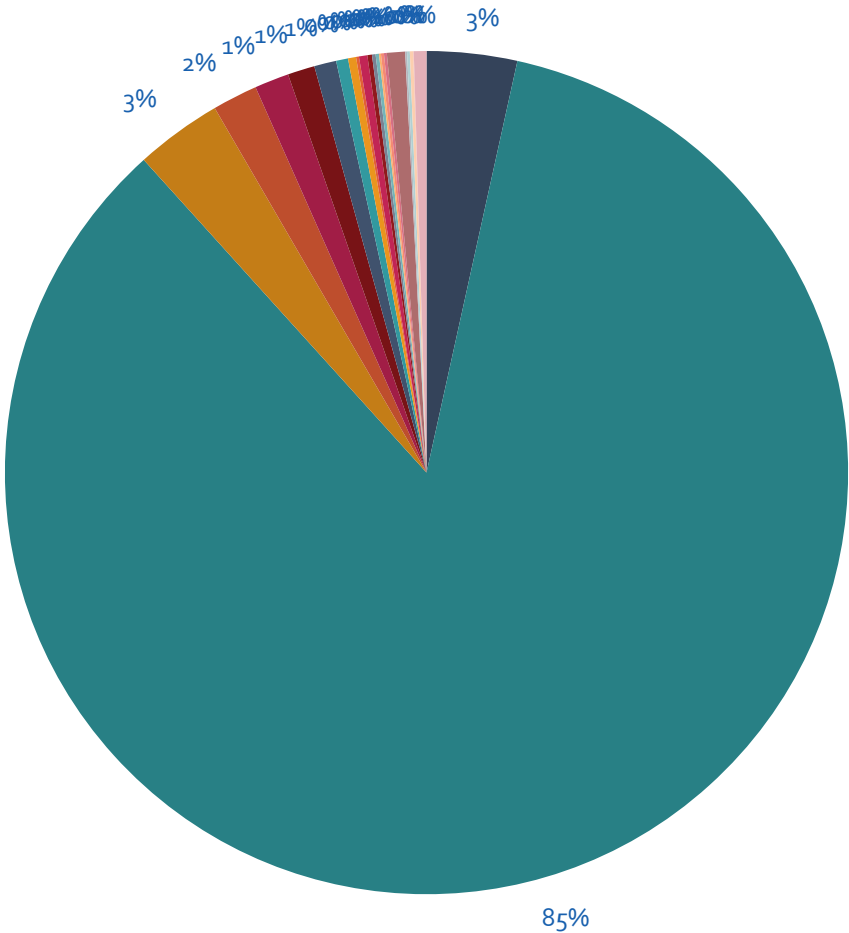
## Un groupe de travail pour formaliser les méthodes?

- ◆ [rossi@ird.fr](mailto:rossi@ird.fr)

# Statistiques de consultation : «Human access», robots ou spam ?

- ◊ 154.126.12.202.MG - - [12/Feb/2017:17:50:33 +0100] "GET /exl-doc/pleins\_textes/pleins\_textes\_5/b\_fdi\_08-09/10230.pdf HTTP/1.1" 200 2300612  
"http://www.google.fr/search?ei=bZGgWJHPllzhvATnxpywDQ&q=trachysphaera+fructigena+pdf&btnG=" "Mozilla/5.0 (Linux; U; Android 4.2.2; en-US; HTC Butterfly Build/JDQ39) AppleWebKit/534.30 (KHTML, like Gecko) Version/4.0 UCBrowser/11.1.5.890 U3/0.8.0 Mobile Safari/534.30"
- ◊ 66.249.64.221.US - - [07/Feb/2017:12:01:04 +0100] "GET /exl-doc/pleins\_textes/divers17-01/010068365.pdf HTTP/1.1" 200 631342 "-" "Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)"

# Referer : merci Google ?



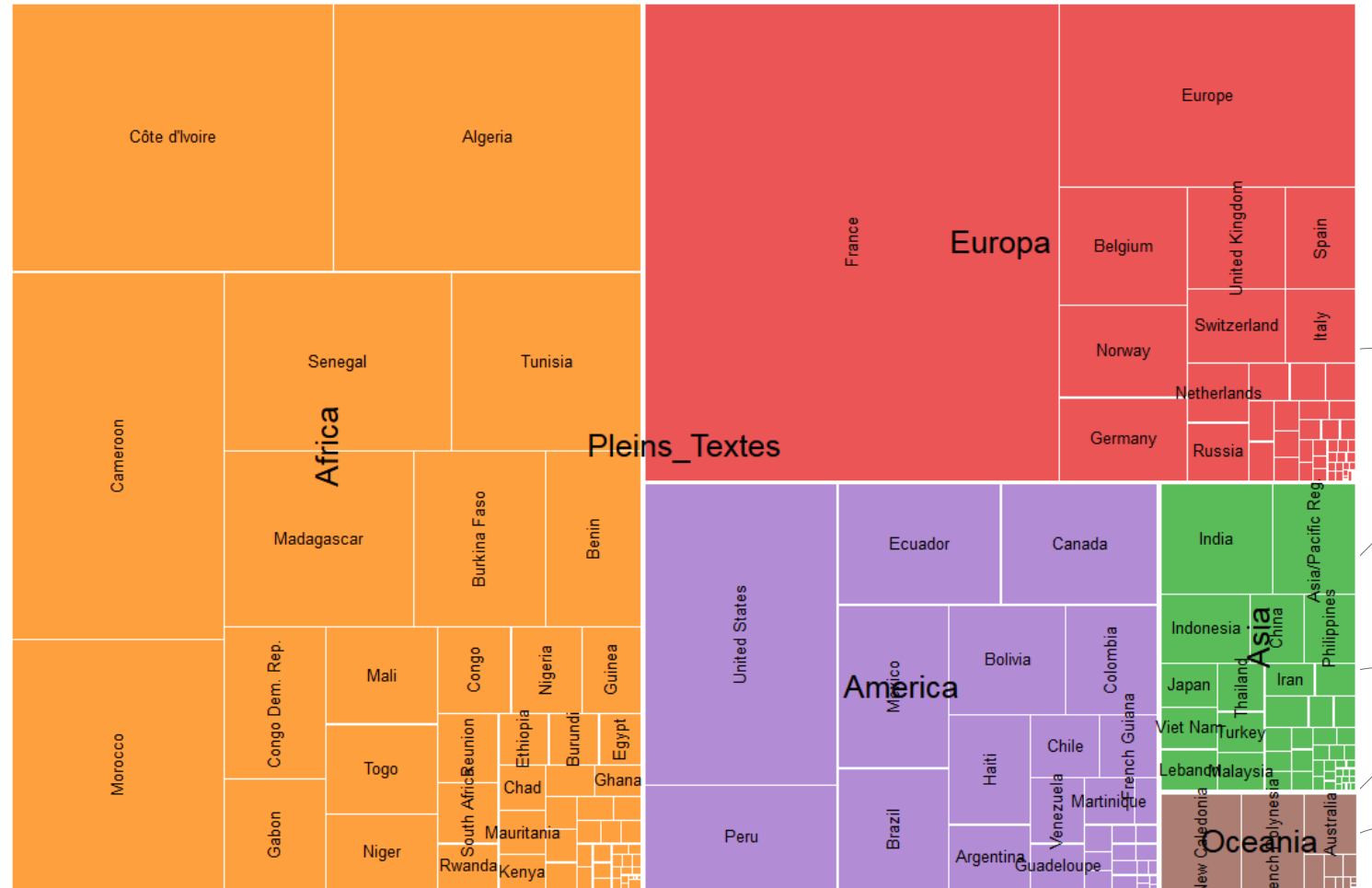
- scholar
- google
- horizon.documentation
- www.documentation
- bing
- myway
- yahoo
- "-pdf"
- ask.com
- yandex
- facebook
- wikipedia
- qwant
- livrespourtous
- baidu
- lexilogos
- ecosia
- .com
- .net



# Une archive ouverte aux Suds ?

## Données 2017

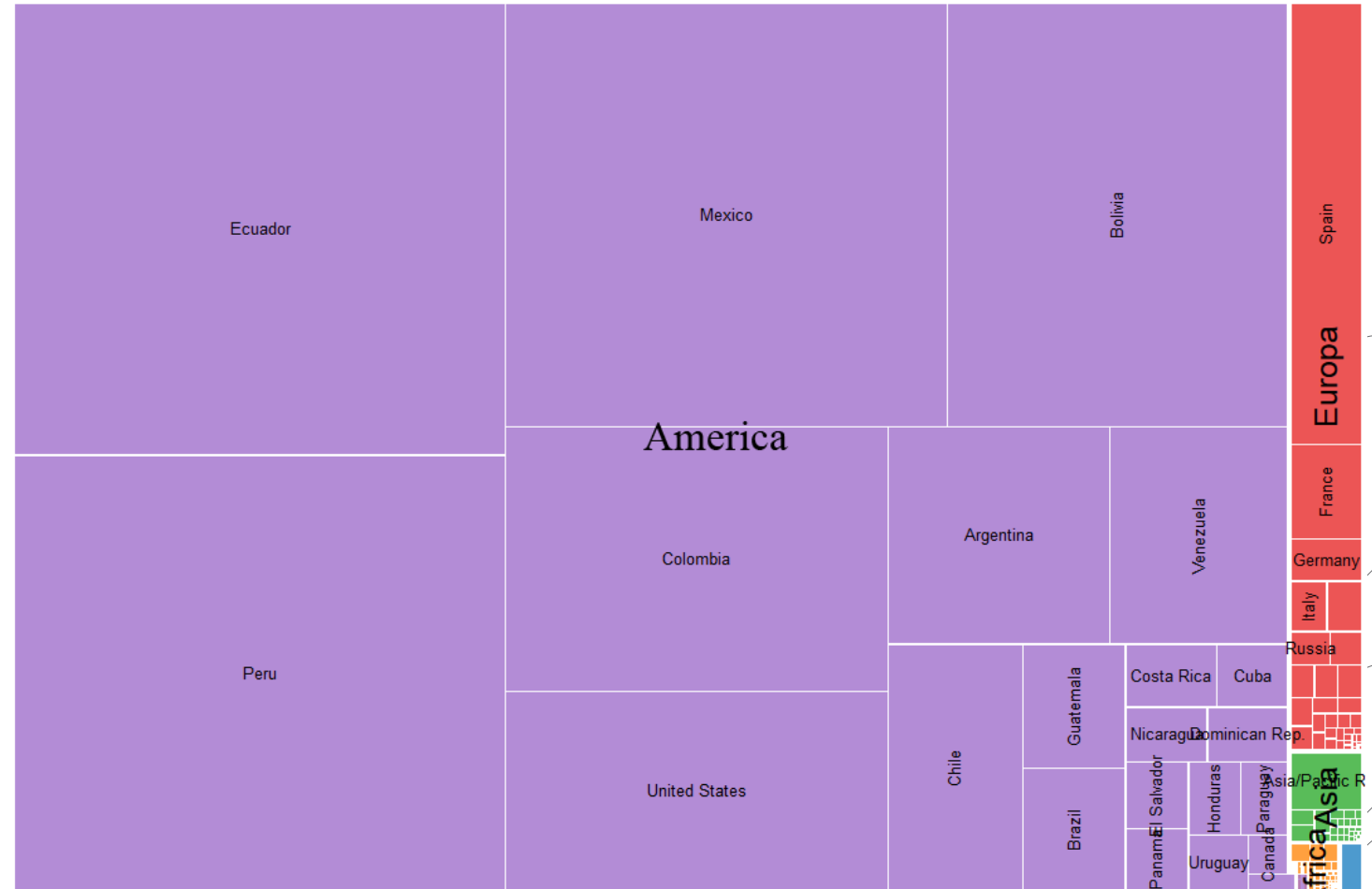
- ◆ 2.749.071 human accesses, 234 pays, 47% Afrique, 65 % Suds
- ◆ Nombre de fichiers "Pleins\_Textes" consultés : 56.761
- ◆ Facteur D : 651
- ◆ Moyenne de consultations : 48.4



# Une archive ouverte aux Suds ?

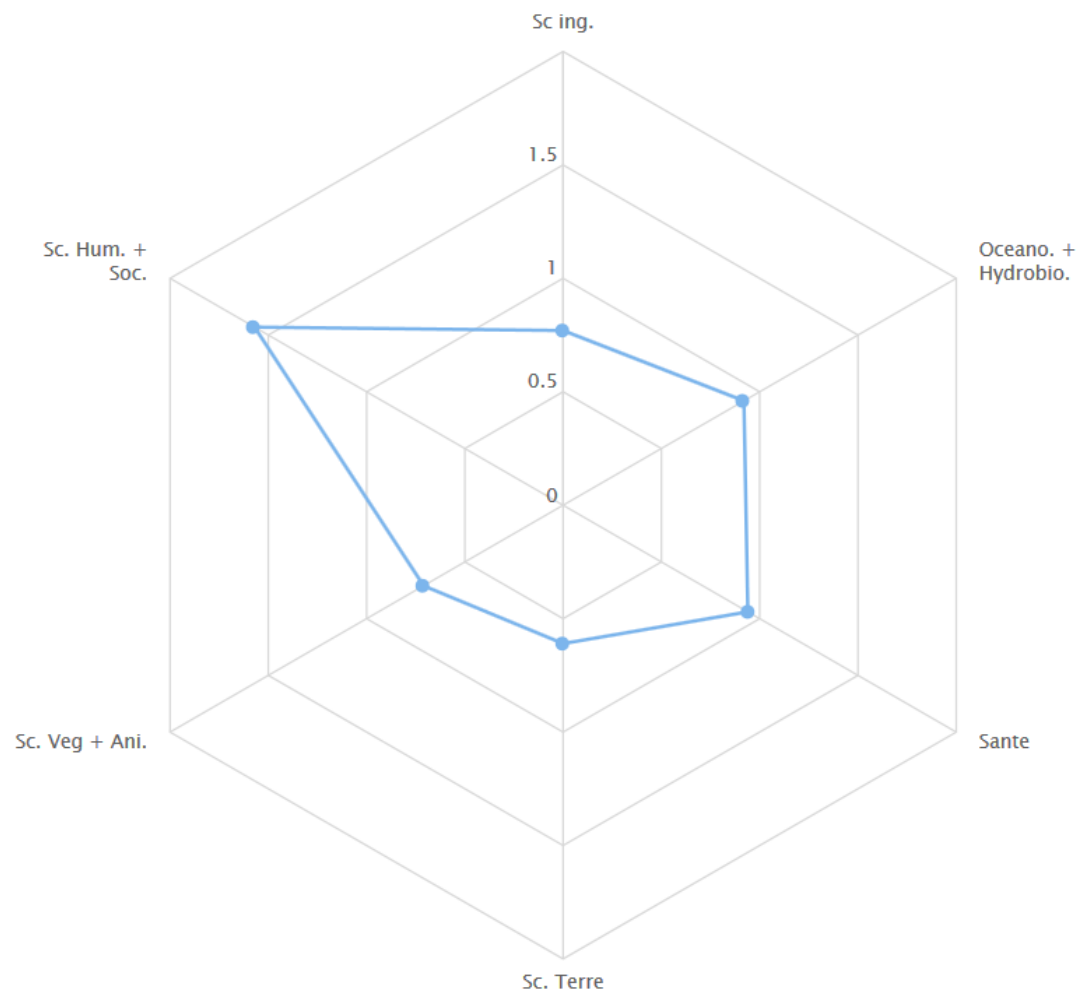
Données 2017,  
documents en espagnol

- ◇ 234.059 human accesses, 161 pays, 94% America, 88% A. latine
- ◇ Nombre de fichiers "Pleins\_Textes" consultés : 2.241
- ◇ Facteur D : 220
- ◇ Moyenne de consultations : 104.4



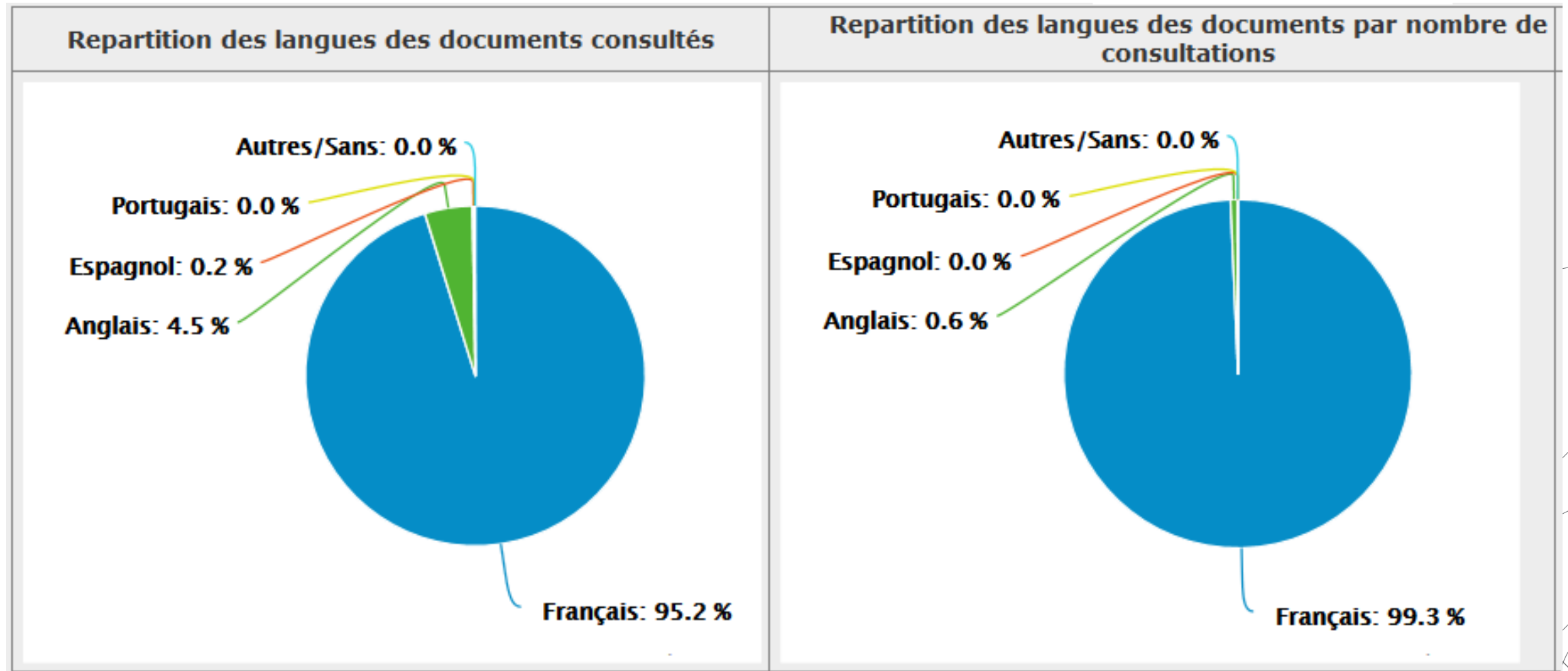
# Cote d'Ivoire : un intérêt pour les Sciences humaines et sociales ?

- ◆ 199.424 human accesses
- ◆ Nombre de fichiers "Pleins\_Textes" consultés : 16.408
- ◆ Facteur D : 175
- ◆ Moyenne de consultations : 12.2





# Cote d'Ivoire : dans quelle langue publier ?



# Les statistiques de consultation : un champ de recherche ?

- ◆ Géolocalisation des consultations, des indexations, des spams, des équipements, ...
- ◆ Montée en charge des consultations en fonction du temps, des thématiques, des langues, ...
- ◆ Définition de nouveaux indicateurs : « downloading factor » « indice d'appropriation », distribution « Pareto », ...
- ◆ Analyse des modes de consultation en fonction des pays
- ◆ Analyse des questions posées (Text mining)



# Des tableaux de bord dynamiques ?

- ◆ **Documents en français :**  
[horizon.documentation.ird.fr/exl-doc/pleins\\_textes/acces/FR-2017.html](http://horizon.documentation.ird.fr/exl-doc/pleins_textes/acces/FR-2017.html)
- ◆ **Documents en anglais:** [horizon.documentation.ird.fr/exl-doc/pleins\\_textes/acces/EN-2017.html](http://horizon.documentation.ird.fr/exl-doc/pleins_textes/acces/EN-2017.html)
- ◆ **Documents en espagnol:**  
[horizon.documentation.ird.fr/exl-doc/pleins\\_textes/acces/ES-2017.html](http://horizon.documentation.ird.fr/exl-doc/pleins_textes/acces/ES-2017.html)
- ◆ **Documents en portugais:**  
[horizon.documentation.ird.fr/exl-doc/pleins\\_textes/acces/PO-2017.html](http://horizon.documentation.ird.fr/exl-doc/pleins_textes/acces/PO-2017.html)

**Merci pour votre attention !**