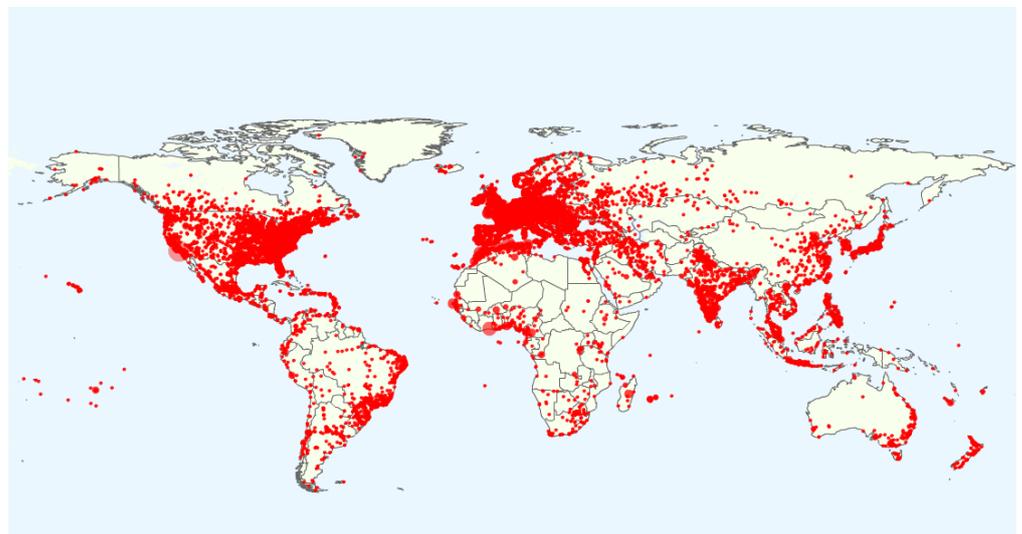


# Statistiques d'usage des documents accessibles en Libre Accès via Archimer

Période 2005-2014



## Table des matières

---

1. Introduction .....	2
2. Statistiques d'utilisation.....	3
2.1. Téléchargements.....	3
2.2. Chemin d'accès .....	10
2.3. Internautes.....	11
3. Dépôts .....	16
4. Méthodologie.....	18

### 1. Introduction

---

Ce document propose un ensemble de données statistiques sur l'usage des documents déposés en Libre Accès dans Archimer, l'Archive Institutionnelle de l'Ifremer.

Seuls les téléchargements du texte intégral des documents ont été pris en compte. La visualisation des fiches de métadonnées n'a pas été comptabilisée.

Pour éviter que l'activité d'administration de l'Archive (dépôt, contrôle, validation ...) ne fausse les résultats, les téléchargements réalisés depuis le réseau informatique de l'Ifremer n'ont pas été pris en compte.

Pour plusieurs données statistiques, des focus sur trois types de documents spécifiques sont proposés : les publications indexées dans le Web Of Sciences® (WOS), les thèses et les rapports<sup>1</sup>. Les publications indexées dans le WOS<sup>2</sup> et les thèses sont particulièrement intéressantes du fait de leur homogénéité en termes de taille et de langue notamment.

Malgré le soin apporté au filtrage du passage des robots, les chiffres présentés contiennent sans doute encore un faible pourcentage d'activités non significatives. Ils doivent donc être considérés comme un ordre de grandeur de bonne qualité et non pas comme des valeurs parfaitement exactes.

---

<sup>1</sup> Rapport de contrats, rapport scientifique, compte-rendu (d'essais, de campagnes, de métrologie...), note (de synthèses, de veille prospective ou technologique...) ...

<sup>2</sup> Nous avons différencié les publications indexées dans le WOS du reste des articles disponibles dans Archimer car elles sont souvent accessibles gratuitement via Archimer mais aussi, sous réserve d'abonnement, via le site de leurs éditeurs (contrairement, par exemple, aux articles de plusieurs journaux anciennement publiés par l'Ifremer, qui sont exclusivement accessibles via Archimer)

## 2. Statistiques d'utilisation

### 2.1. Téléchargements

Figure 1 : Progression de la somme annuelle des téléchargements des documents<sup>3</sup>

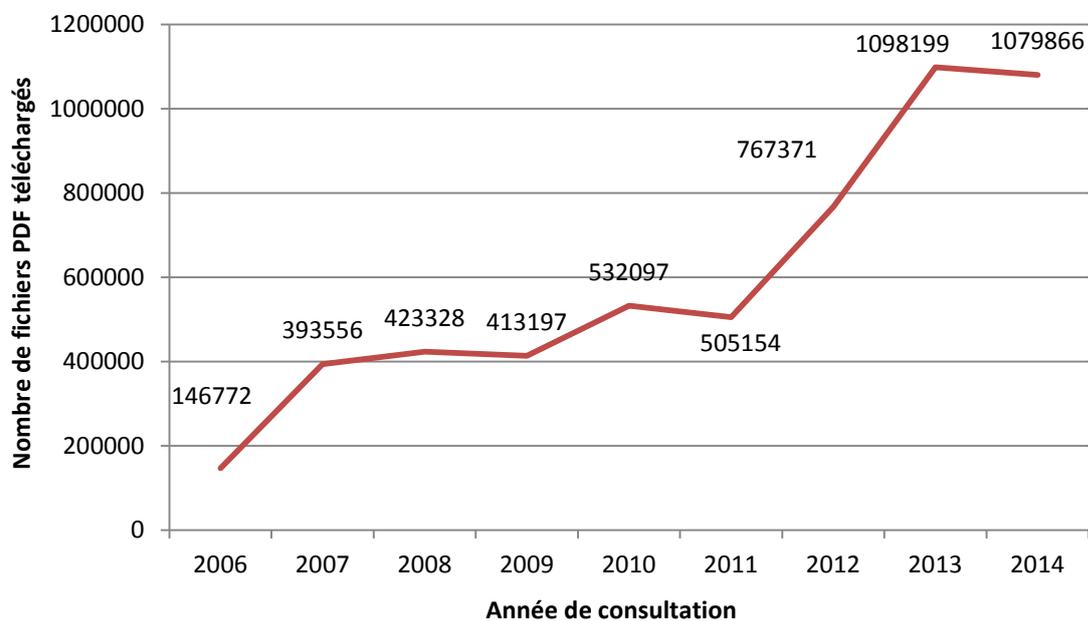
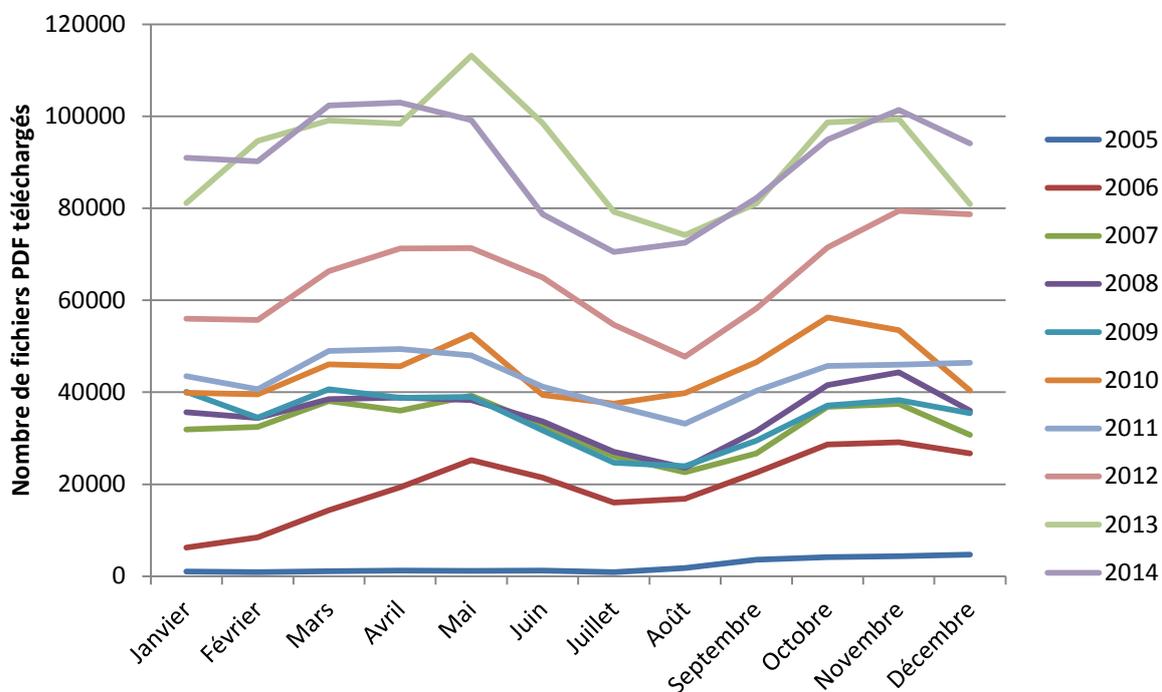
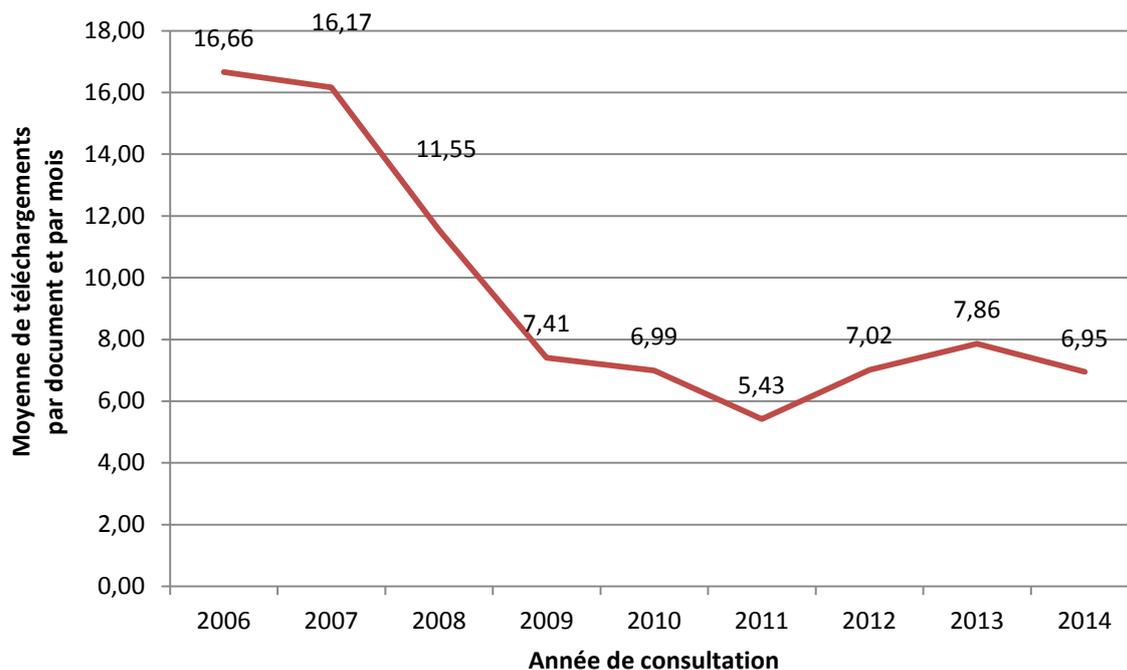


Figure 2 : Evolution de la somme mensuelle des téléchargements des documents



<sup>3</sup> Tout type de documents confondus : publication, acte de colloque, rapport, thèse, HDR, avis, expertise, ouvrage, chapitre d'ouvrage, poster.

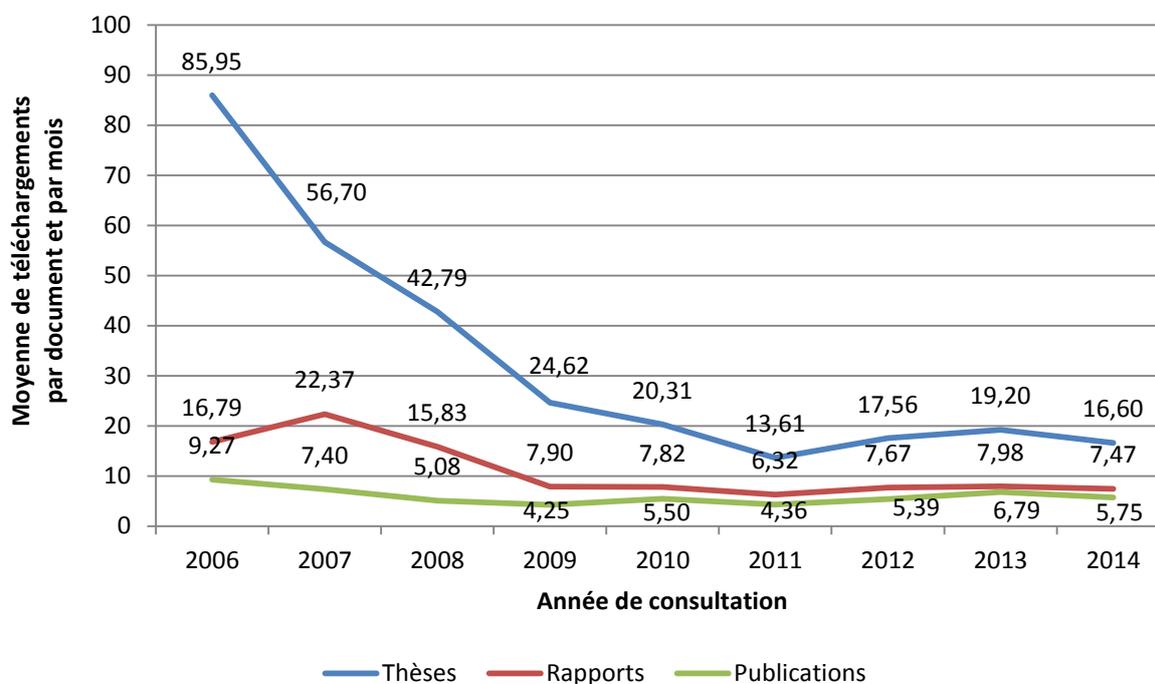
Figure 3 : Evolution de la moyenne mensuelle des téléchargements par document



Remarque : les moyennes mensuelles de téléchargement lors d'une année  $n$  sont calculées ainsi :

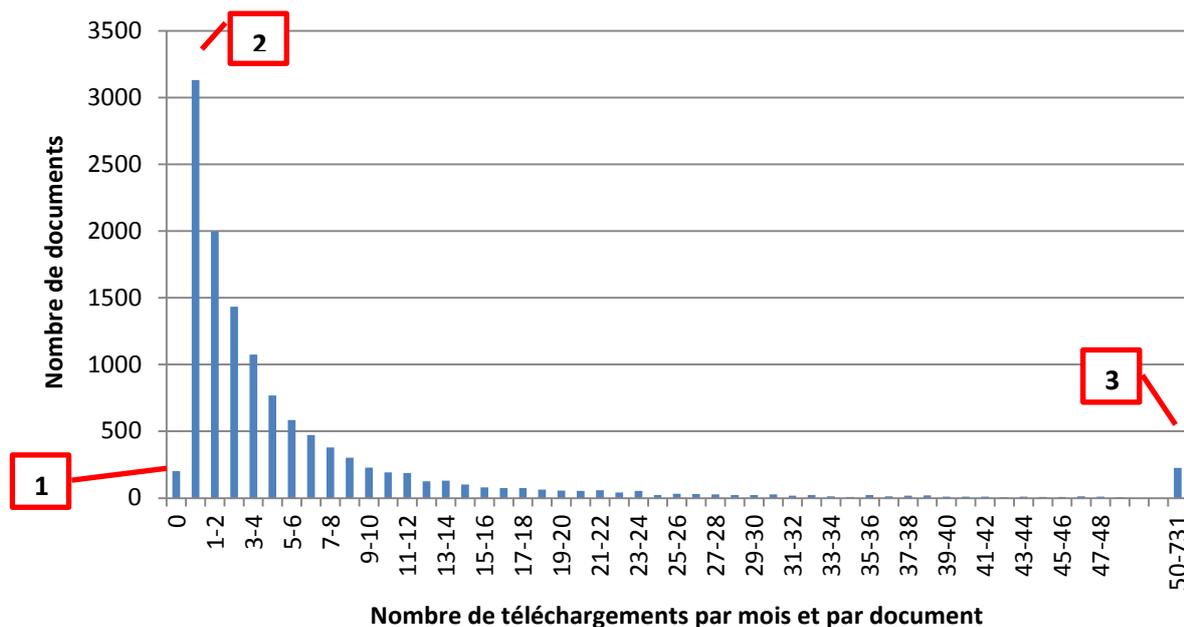
$$\frac{\text{Téléchargements enregistrés lors de l'année } n \text{ des documents déposés avant la fin de l'année } n - 1}{(\text{Nombre de documents déposés avant la fin de l'année } n - 1) \times 12}$$

Figure 4 : Evolution de la moyenne mensuelle des téléchargements des publications<sup>4</sup>, des thèses et des rapports



<sup>4</sup> Publications indexées dans le WOS

Figure 5 : Histogramme des moyennes mensuelles de téléchargements des documents au cours de l'année 2014



La Figure 5 se lit de la façon suivante :

1. 202 des 12526 documents déposés en Libre Accès avant le 31 décembre 2013 dans Archimer n'ont jamais été téléchargés au cours de l'année 2014
2. 3130 de ces 12526 documents ont été téléchargés de 1 à 11 fois au cours de l'année 2014. La moyenne mensuelle de téléchargements de ces 3130 documents se situe donc entre 0 et 1.
3. 227 de ces 12526 documents ont été téléchargés en moyenne entre 50 et 731 fois par mois au cours de l'année 2014.

Figure 6 : Histogramme des moyennes mensuelles de téléchargements des publications au cours de l'année 2014

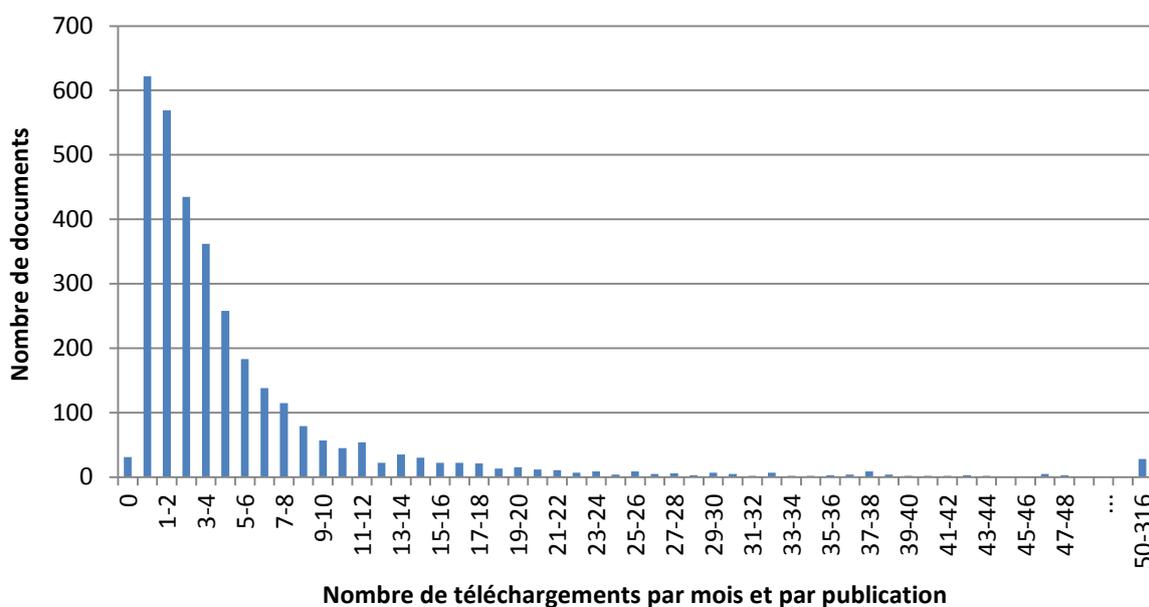


Figure 7 : Histogramme des moyennes mensuelles de téléchargements des thèses au cours de l'année 2014

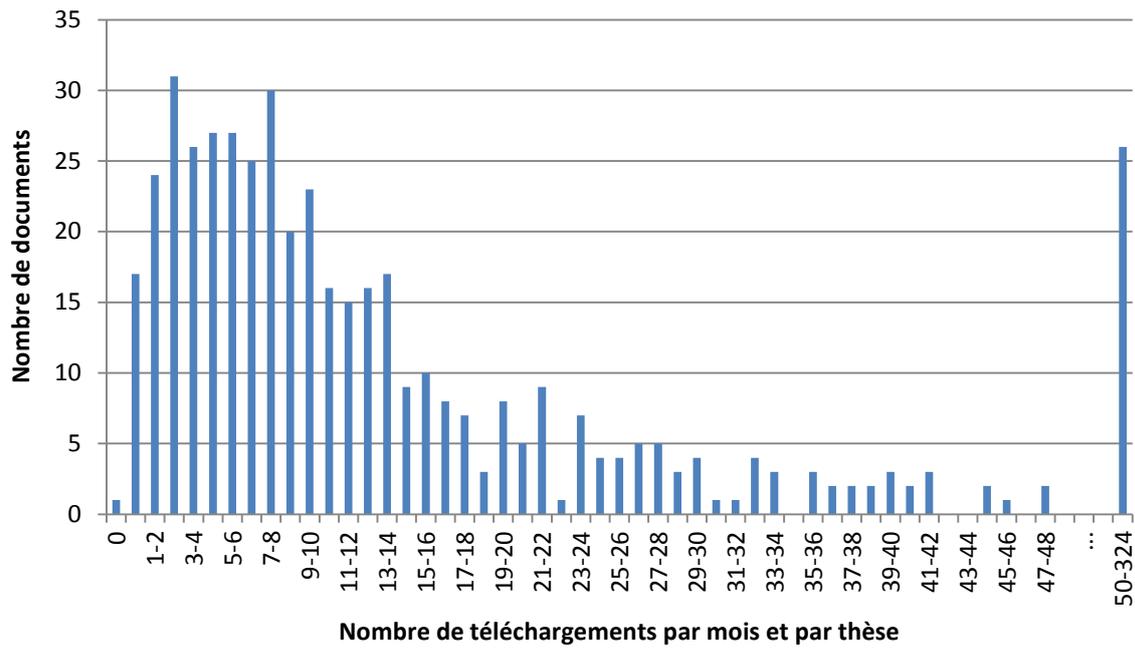


Figure 8 : Histogramme des moyennes mensuelles de téléchargements des rapports au cours de l'année 2014

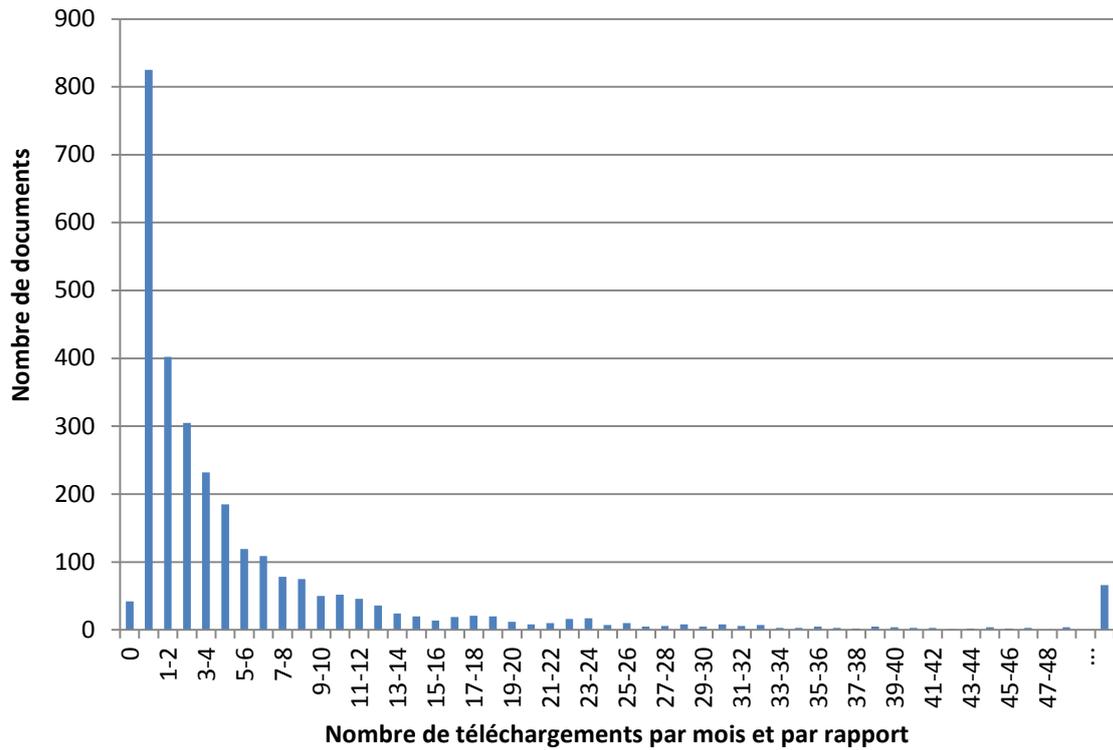


Figure 9 : Evolution de la moyenne mensuelle des téléchargements des thèses par année de dépôt

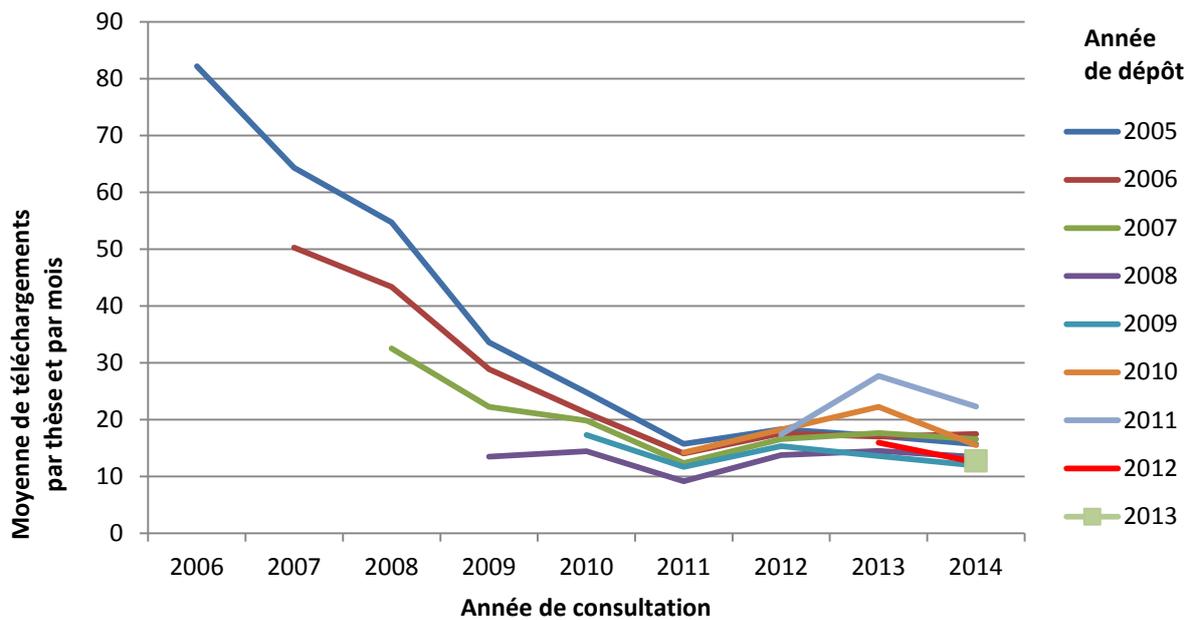


Figure 10 : Evolution de la moyenne mensuelle des téléchargements des publications par année de dépôt

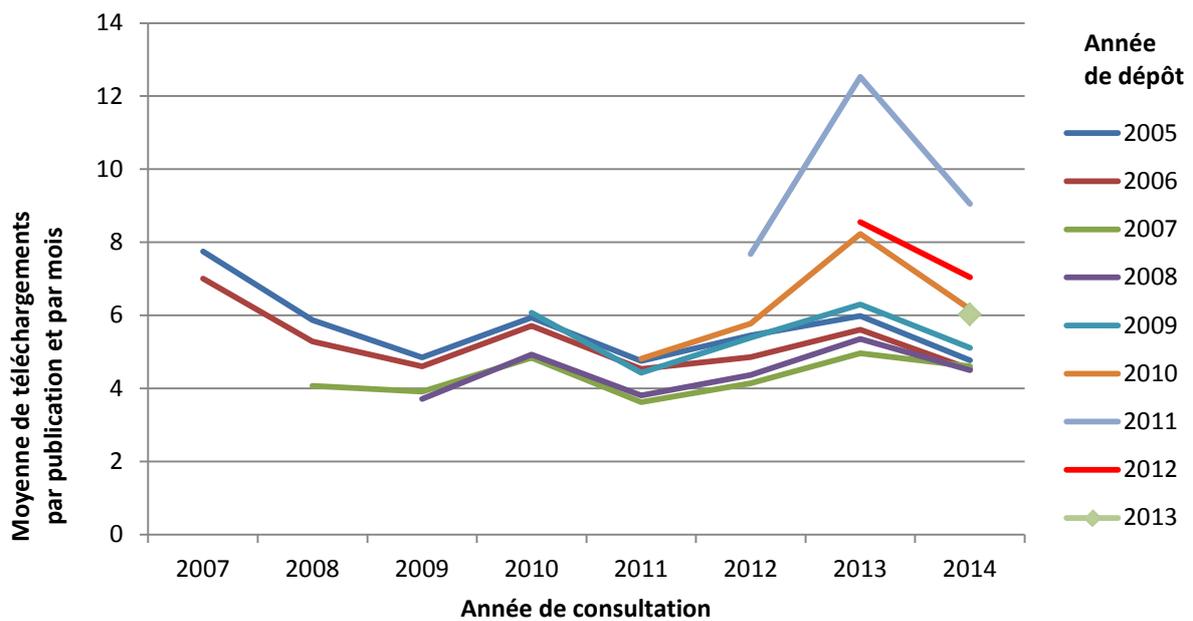


Figure 11 : Evolution de la moyenne mensuelle des téléchargements des rapports par année de dépôt

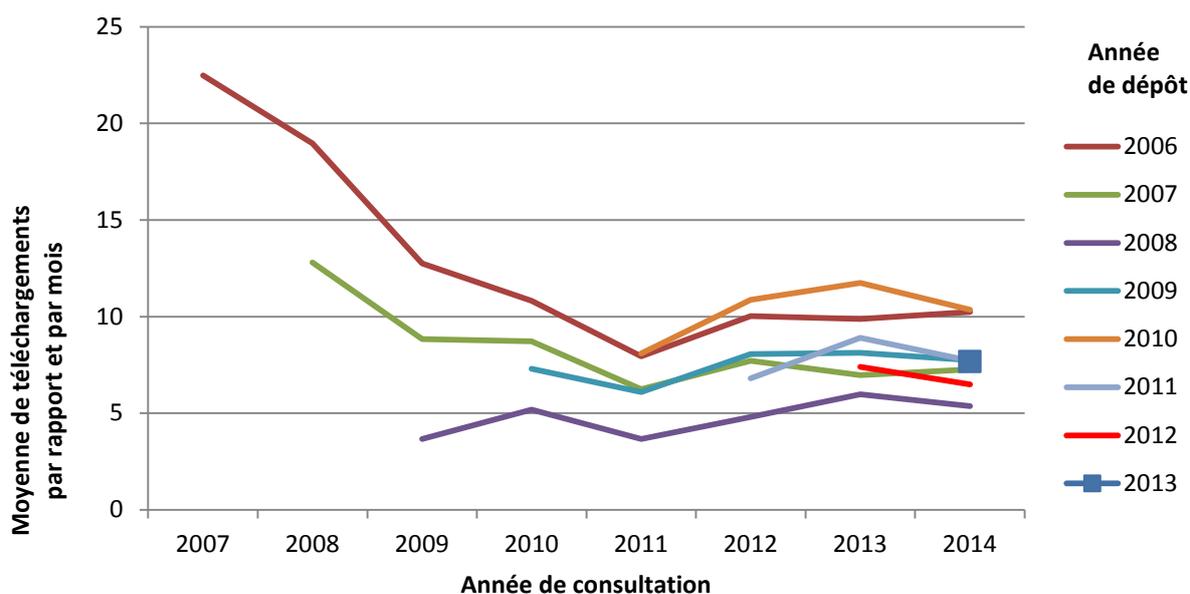
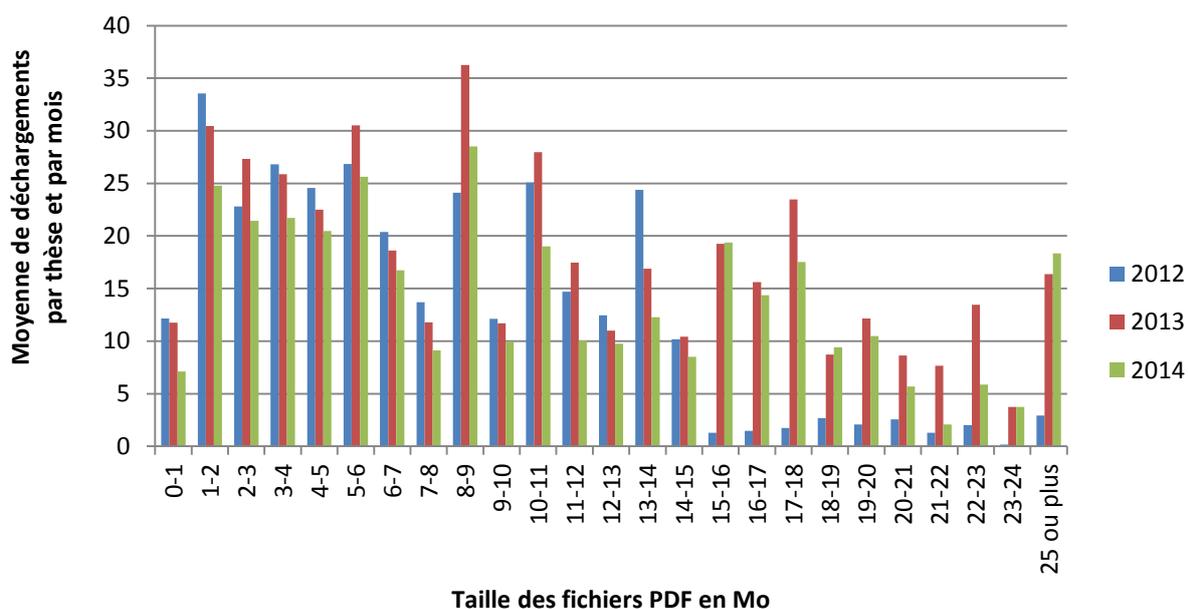
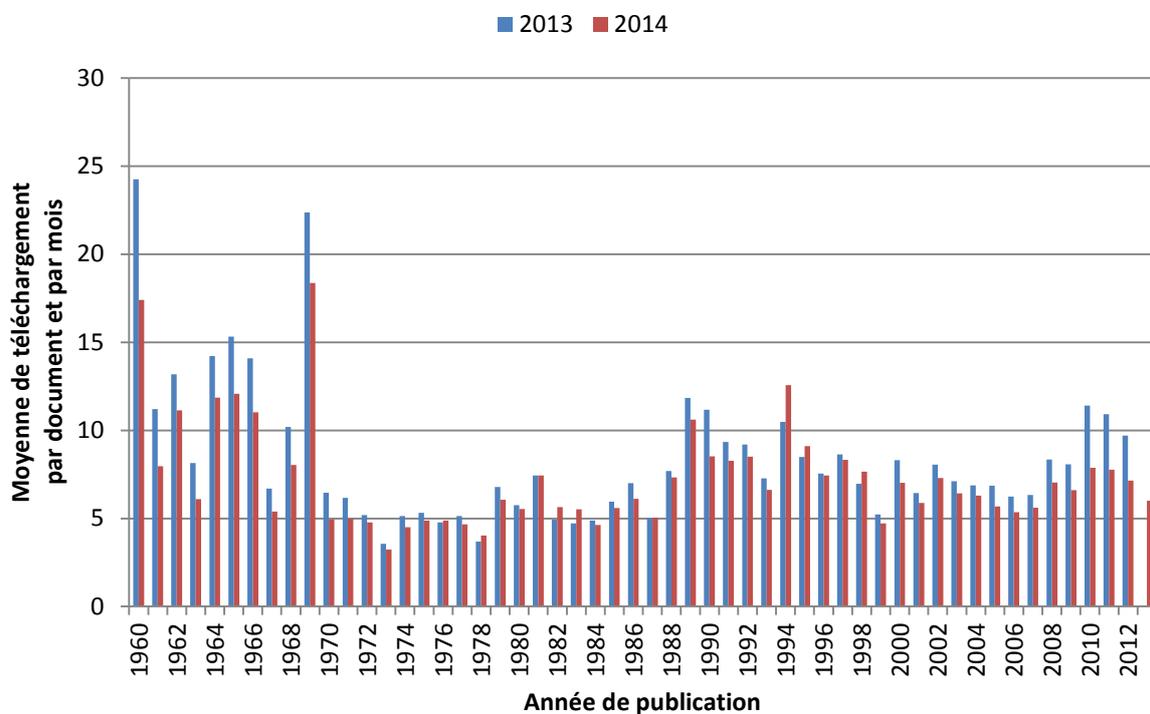


Figure 12 : Histogramme des moyennes mensuelles de téléchargements des thèses en fonction de la taille des fichiers



Remarque : La différence très marquée entre les téléchargements des fichiers de tailles inférieures et supérieures à 15 Mo a disparu en 2013. Elle était sans doute liée au fait que Google ne semblait pas indexer les fichiers volumineux (la limite semblait se situer autour de 10Mo à 15 Mo en fonction des années). Seules les fiches de métadonnées des gros fichiers étaient indexées. Cette limite semble avoir évolué en 2013 puisqu'il est désormais possible de trouver dans les index de Google des fichiers d'une taille supérieure à 60 Mo.

Figure 13 : Histogramme des moyennes mensuelles de téléchargements des documents en fonction de leur année de publication au cours de l'année 2013 (bleu) et 2014 (rouge)



## 2.2. Chemin d'accès

Tableau 1 : Sources d'accès aux documents téléchargés au cours du mois de Novembre 2014

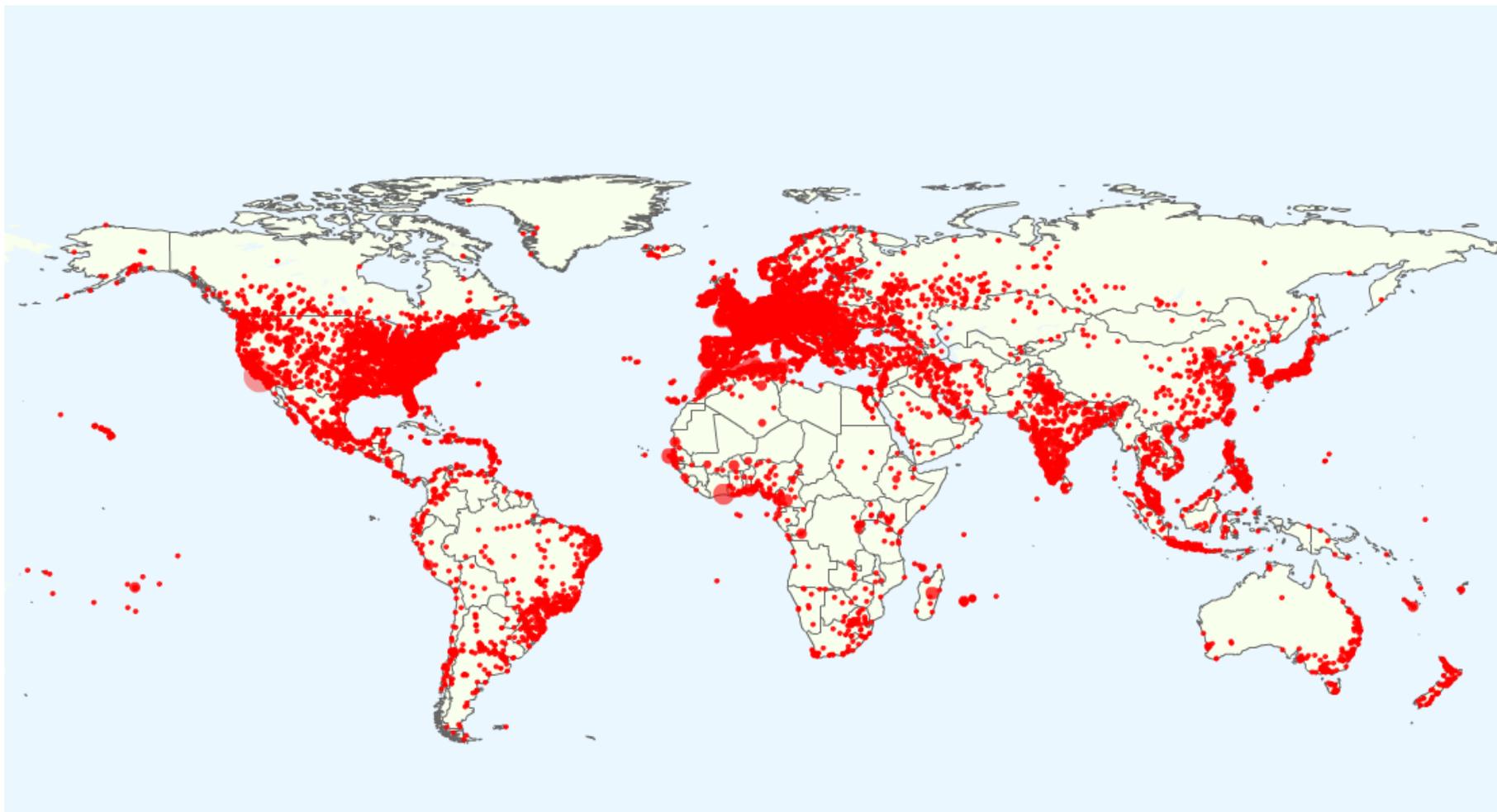
Source des téléchargements	Nombre de téléchargements	Pourcentage des téléchargements <sup>5</sup>
Google	74408	84,28%
Google Scholar	5620	6,37%
Archimer	1207	1,37%
Ask	902	1,02%
Bing	728	0,82%
Baidu	652	0,74%
Yahoo	619	0,70%
Annuaire Ifremer	303	0,34%
Facebook	241	0,27%
Sites Ifremer	238	0,27%
Wikipedia	216	0,24%
Isidore	201	0,23%
Sogou	192	0,22%
Worldcat	140	0,16%
Yandex	98	0,11%
... <sup>6</sup>		
Twitter	63	0,07%
Eaufrance	55	0,06%
Gallica	50	0,06%
ResearchGate	9	0,01%

<sup>5</sup> Pourcentage par rapport aux requêtes dont la source est identifiable via la Referring URL

<sup>6</sup> Ce tableau ne liste pas un grand nombre de systèmes qui pourraient s'apparenter à du SPAM. Certains systèmes réindexent le contenu de documents disponibles dans Archimer et les diffusent sous un format différent (ex : banque-pdf.fr ...). D'autres s'apparentent à des barres de recherche qui interceptent les recherches des internautes (ex : babylon, conduit ...) avant de les rediriger vers Google.

### 2.3. Internautes

Figure 14 : Localisation des téléchargements enregistrés au cours de l'année 2014



Remarque : 18% des téléchargements 2013 ne sont pas localisés sur cette carte. Un point peut représenter jusqu'à 32 000 téléchargements.

Tableau 2 : Liste des principaux pays utilisateurs des documents rédigés en anglais

<b>Pays</b>	<b>Nombre de téléchargements</b>	<b>Pourcentage des téléchargements</b>
United States	68005	18,3%
France	26647	7,2%
India	26428	7,1%
United Kingdom	23388	6,3%
China	10528	2,8%
Canada	9592	2,6%
Australia	9444	2,5%
Germany	9206	2,5%
Italy	8650	2,3%
Spain	8629	2,3%
Indonesia	8575	2,3%
Ukraine	8475	2,3%
Malaysia	6835	1,8%
Philippines	6814	1,8%
Iran, Islamic Republic of	6366	1,7%
Brazil	6153	1,7%
Japan	5860	1,6%
Korea, Republic of	5151	1,4%
Mexico	4794	1,3%
Netherlands	4768	1,3%
Thailand	4540	1,2%
Portugal	4465	1,2%
Turkey	4426	1,2%
Vietnam	4255	1,1%
Russian Federation	4063	1,1%
Egypt	3587	1,0%

Tableau 3 : Liste des principaux pays utilisateurs des documents rédigés en français

<b>Pays</b>	<b>Nombre de téléchargements</b>	<b>Pourcentage des téléchargements</b>
France	323462	45,7%
Algeria	79025	11,2%
Morocco	46213	6,5%
United States	42206	6,0%
Tunisia	35488	5,0%
Canada	16276	2,3%
Cote D'Ivoire	12774	1,8%
Belgium	12157	1,7%
Cameroon	9901	1,4%
Madagascar	8698	1,2%
Senegal	6823	1,0%
Reunion	5749	0,8%
Switzerland	4462	0,6%
China	4237	0,6%
Benin	3435	0,5%
Spain	3378	0,5%
Guadeloupe	3194	0,5%
United Kingdom	3046	0,4%
Martinique	2987	0,4%
French Polynesia	2466	0,3%
New Caledonia	2324	0,3%
Congo, The Democratic Republic of the	2307	0,3%
Germany	2262	0,3%
Burkina Faso	2145	0,3%

Tableau 4 : Liste des principaux utilisateurs académiques et privés des documents rédigés en anglais au cours du mois de Novembre 2014

Institution	Pays	Nombre de téléchargements
National Knowledge Network	India	5073
Renater	France	2654
Research Organization of Information and Systems	Japan	601
Universitaet Bremen	Germany	557
Research network University of Ghent	Belgium	546
Universitaet Kiel	Germany	503
Consiglio Nazionale delle Ricerche	Italy	501
Wageningen University and Research Centre	Netherlands	479
University of Southampton	United Kingdom	431
Universiti Malaysia Terengganu	Malaysia	427
Croatian Academic and Research Network	Croatia	426
The University of Plymouth	United Kingdom	424
Dalhousie University	Canada	395
Commissariat a l'Energie Atomique	France	389
Universidade do Porto	Portugal	386
Centre de Ressources Informatiques, Universite de	France	379
University of California, San Diego	United States	376
Universiti Putra Malaysia	Malaysia	362
Universite de Bordeaux, Departement TIC	France	331
Navy Network Information Center (NNIC)	United States	321
Malaysian Research & Education Network	Malaysia	312
Universidade de Aveiro	Portugal	311
TOTAL S.A.	France	304
University of Wales Bangor	United Kingdom	298
Woods Hole Oceanographic Institution	United States	296
Universite Pierre et Marie Curie	France	294
Bogor Agricultural University	Indonesia	291
Universidade de Sao Paulo	Brazil	279
University of Washington	United States	278
UBO - Universite de Brest	France	254
Universitas Brawijaya	Indonesia	240
University of Aberdeen, UK	United Kingdom	235
Universite de La Rochelle	France	234
University of Tasmania	Australia	233
Imperial College London	United Kingdom	231
National Taiwan University	Taiwan	230
CRI Universite de Nantes - 2 Rue de la Houssiniere	France	229
Universite Montpellier II	France	229
University of Stirling	United Kingdom	228

Tableau 5 : Liste des principaux utilisateurs académiques et privés des documents rédigés en français au cours du mois de Novembre 2013

Institution	Pays	Nombre de téléchargements
Algerian Academic Research Network	Algeria	1552
Universite de Caen	France	1383
Algerian Academic Research Network	Algeria	1376
Universite de Bordeaux	France	1288
Universite de Brest	France	1287
Universite de Nantes	France	1189
Universite Montpellier II	France	930
Universite de La Rochelle	France	841
Houari Boumediene's University of Science & Techno	Algeria	817
Université de Rennes	France	577
Electricite de France	France	556
Universite Pierre et Marie Curie	France	539
CNRST (Centre National pour la Recherche Scientifi	Morocco	528
Reseau de la Recherche a Marseille	France	484
University of Littoral - Dunkerque	France	430
Universite Laval	Canada	418
Universite de Liege (ULg)	Belgium	396
Universite du Quebec a Rimouski	Canada	391
Universite de PERPIGNAN	France	367
TOTAL S.A.	France	363
Universite de Rouen	France	361
Universite de Strasbourg	France	355
Universite Cadi Ayyad - Marrekech -	Morocco	353
Gaz de France	France	350
Centre de Calcul el-Khawarizmi - CCK	Tunisia	346
Renater - Banyuls-sur-mer	France	304
Universite de Nice Sophia Antipolis	France	300
Institut Mediteraneen de Technologie	France	299
Research In Motion UK Limited	United Kingdom	299
Renater - Vitrolles	France	295
Universite de Lorraine	France	293
University Mhamed Bouguerra	Algeria	281
Universite des Sciences et Technologies de Lille	France	280
Universite du Quebec a Montreal	Canada	279
CNRST (Centre National pour la Recherche Scientifi	Morocco	273
BRGM	France	271
Universite de la Reunion	Reunion	268
Universite du Sud Toulon - Var	France	242
Renater - Nancy	France	227
Universite Paul Sabatier	France	224

### 3. Dépôts

Figure 15 : Progression du nombre de documents disponibles dans Archimer

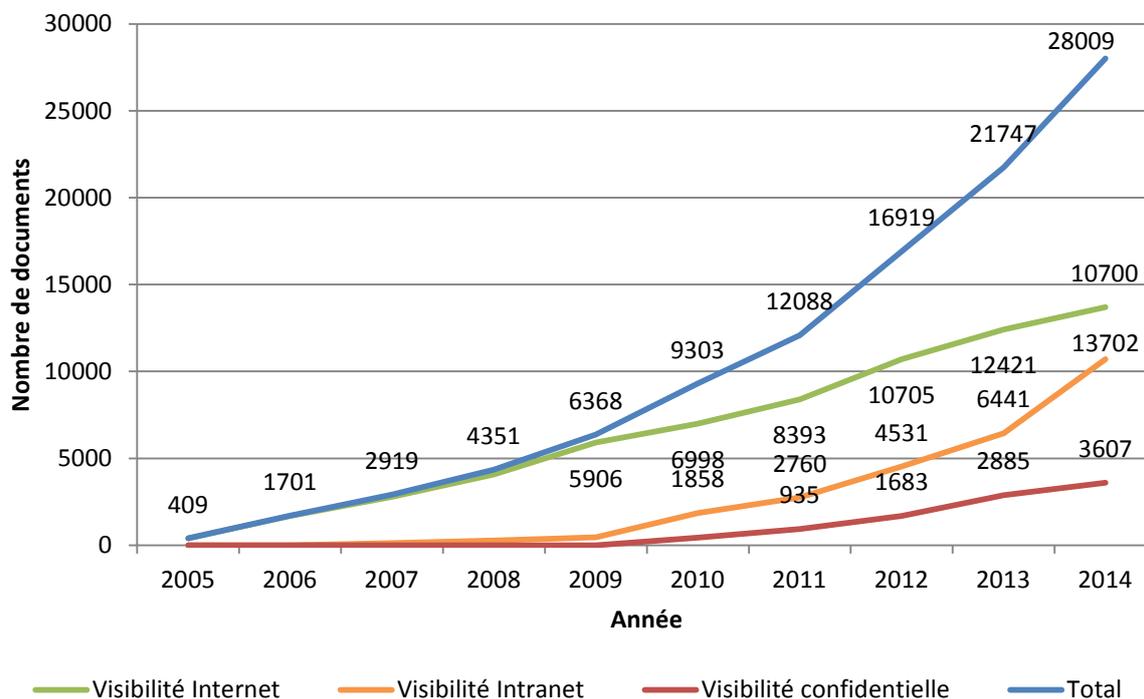
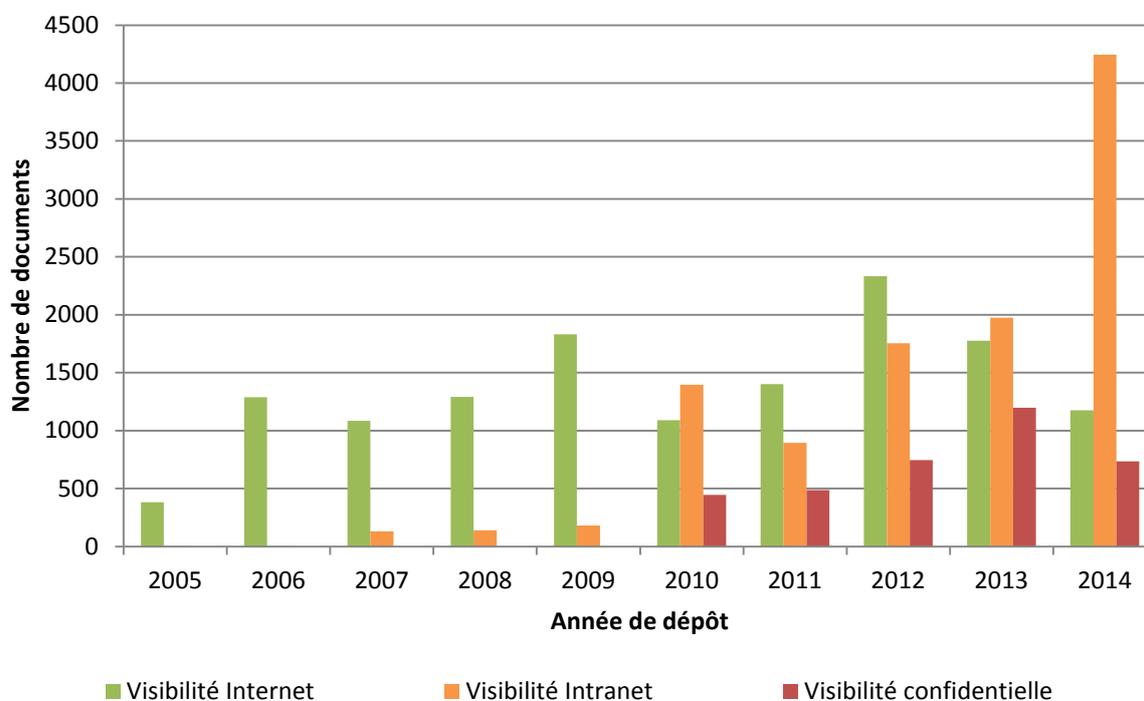
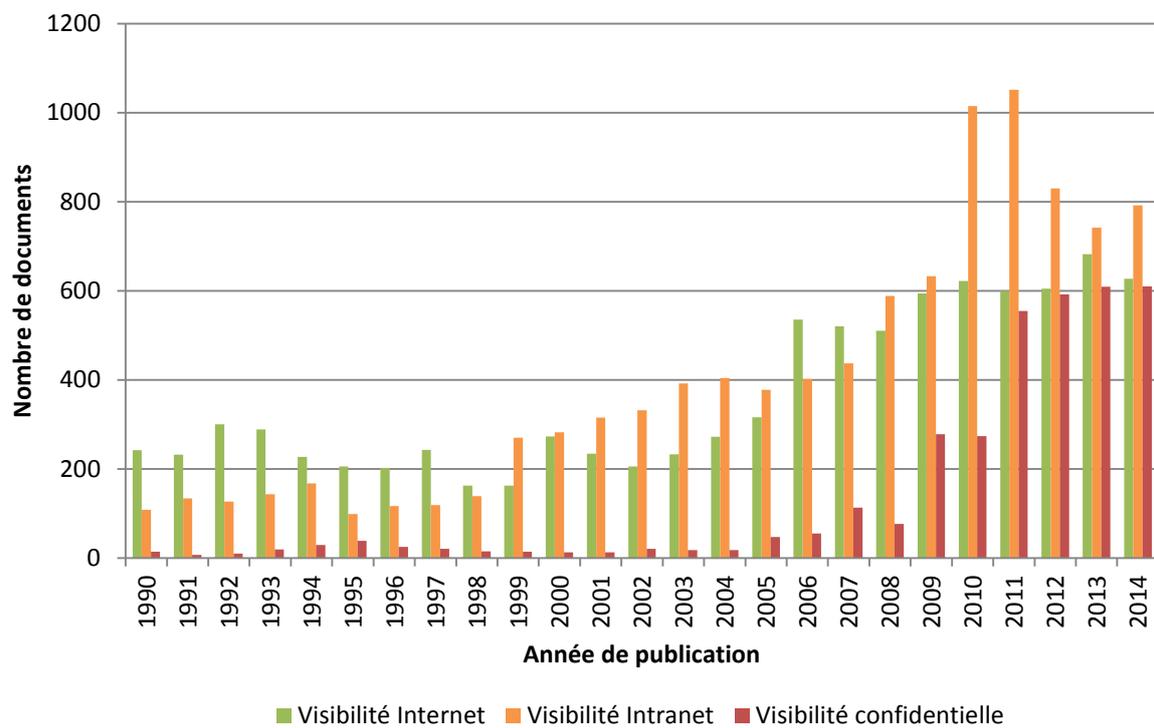


Figure 16 : Histogramme du nombre de documents déposés dans Archimer par année de dépôt



Remarque : le pic de dépôts en 2014 avec une visibilité Intranet correspond au travail de collecte des publications associées aux campagnes à la mer.

Figure 17 : Histogramme des années de publication des documents déposés dans Archimer



## 4. Méthodologie

Les résultats présentés dans ce document sont issus de l'analyse des logs du serveur Web Apache de l'Ifremer. Toutes les nuits, un script informatique repère dans le fichier de logs, les lignes correspondantes aux téléchargements des fichiers PDF disponibles dans Archimer. Ces lignes se présentent sous la forme suivante :

```
xxx.xxx.xxx.xxx - - [22/May/2013:00:28:09 +0200] "GET /doc/00038/14916/12245.pdf HTTP/1.1" 200 409200
"http://scholar.google.it/scholar?q=phytoplankton+diversity+during+the+spring+bloom+in+the+northweste
rn+mediterranean+sea&hl=it&as_sdt=0&as_vis=1&oi=scholar&sa=X&ei=6_SbUcilKsTR7AbGxYGgCQ&sqi=2&v
ed=0CCwQgQMwAA" "Mozilla/5.0 (Windows NT 6.1) AppleWebKit/537.31 (KHTML, like Gecko)
Chrome/26.0.1410.64 Safari/537.31"
```

Les informations ci-dessous sont extraites de chaque ligne :

- **L'adresse IP** (ex : xxx.xxx.xxx.xxx) permet d'obtenir des informations (FAI et coordonnées géographiques) sur l'identité de l'Internaute à l'aide de la base GeolIP<sup>7</sup>.
- **La date et l'heure du téléchargement** (ex : 22/May/2013:00:28:09).
- **L'URL du document téléchargé** (ex : /doc/00038/14916/12245.pdf) contient le numéro d'identification du document dans Archimer (ex : 14916).
- **Le code du résultat de la requête HTTP** (ex : 200)
- **La « Referring URL »** (ex : http://scholar.google.it/scholar?q=phytoplankton+diversity+during...) indique l'URL de la page qui est à l'origine du téléchargement. Elle correspond, dans notre exemple, à l'URL de Google Scholar avec, en paramètre, la question posée par l'internaute dans Google Scholar (ex : phytoplankton diversity during the spring bloom in the northwestern mediterranean sea). Si un internaute accède à un document via sa fiche de métadonnées (ex : <http://archimer.ifremer.fr/doc/00038/14916/>), le script remonte dans le fichier de log pour trouver la « Referring URL » de cette première requête. L'information « Referring URL » n'est pas systématiquement disponible. Les robots, notamment, exécutent souvent des requêtes HTTP sans fournir ce paramètre. Son absence peut donc aider à les identifier.
- **Le « User-Agent »** de la requête HTTP (ex : Mozilla/5.0 Windows NT 6.1 ...) fournit des informations sur le navigateur de l'internaute.

Les robots des moteurs de recherche et les aspirateurs de sites sont à l'origine de la très grande majorité des téléchargements. Plusieurs actions permettent de filtrer leurs activités :

- Les requêtes HTTP dont le « User-Agent » contient un des mots<sup>8</sup> associés à des robots (BOT, SPIDER, CRAWLER ...) sont supprimées.
- Les téléchargements provenant de plusieurs FAI (ex : OVH, Microsoft Hosting ...) sont également exclues en l'absence de paramètre « Referring URL ».
- Un grand nombre de téléchargements exécutés à partir de la même adresse IP est systématiquement suspect : les requêtes sont alors vérifiées manuellement (au-delà de 100 requêtes par mois). Dans certains cas, ces téléchargements proviennent d'une université dont les appels sortant sont filtrés par un proxy, mais, dans la majorité des cas, il s'agit de robots dont les téléchargements sont alors supprimés.
- ...

<sup>7</sup> <http://www.maxmind.com/fr/home>

<sup>8</sup> Une liste de robots à exclure est disponible sur le site du projet COUNTER : [http://www.projectcounter.org/code\\_practice.html](http://www.projectcounter.org/code_practice.html)