

# Publications françaises Open Access 2010-2014 Méthodologie

Septembre 2016



|   |           |
|---|-----------|
| <b>Introduction .....</b>   | <b>4</b>  |
| <b>Open Access .....</b>  | <b>4</b>  |
| <b>Constitution du corpus France.....</b>   | <b>5</b>  |
| <b>Repérage et marquage des périodiques OA ou « Hybride » .....</b>                                   | <b>5</b>  |
| <b>Extraction et organisation des données .....</b>   | <b>5</b>  |
| <b>Répercussion du statut du périodique sur les publications .....</b>                                | <b>6</b>  |
| <b>Traitement des périodiques .....</b>   | <b>6</b>  |
| Traitement des titres des périodiques .....   | 7         |
| Marquage des périodiques OA : JCR – DOAJ .....  | 7         |
| Marquage des périodiques OA : ROAD – « Grands Editeurs » .....  | 8         |
| Marquage des périodiques « Hybride » des 7 « Grands Editeurs » .....                                  | 8         |
| Marquage des périodiques non OA des « Grands Editeurs » .....   | 8         |
| Marquage des périodiques sans statut .....  | 8         |
| Résultat du marquage .....  | 8         |
| <b>Marquage des notices OA .....</b>  | <b>11</b> |
| <b>Répercussion du marquage des périodiques OA.....</b>   | <b>11</b> |
| <b>Traitement des notices issues de périodiques « Hybride » ou de périodiques « SansStatut ».....</b> | <b>11</b> |
| <b>Enrichissement des notices OA .....</b>  | <b>13</b> |
| <b>Marquage thématique.....</b>   | <b>13</b> |
| <b>Homogénéisation des titres des périodiques .....</b>   | <b>13</b> |
| <b>Homogénéisation des éditeurs.....</b>  | <b>13</b> |
| <b>Attribution des APCs.....</b>  | <b>14</b> |
| DOAJ .....  | 14        |
| Fichier Inserm .....  | 14        |
| Traitement des APCs.....  | 14        |
| <b>Traitement des champs « Affiliations » et « Reprint Author » .....</b>                             | <b>14</b> |
| Champ « Affiliations » .....  | 14        |
| Champ « Reprint Author » .....  | 15        |
| <b>Vérification des données périodiques .....</b>   | <b>15</b> |
| <b>Constitution du corpus OA final.....</b>   | <b>15</b> |
| Homogénéisation des thématiques .....   | 15        |
| Homogénéisation des titres .....  | 15        |
| Homogénéisation des éditeurs.....   | 15        |
| Homogénéisation des données APC.....  | 15        |
| Format de la notice finale d'une publication .....  | 16        |
| <b>Statistiques .....</b>   | <b>16</b> |
| <b>Liste des tableaux &amp; figures</b>   |           |
| Tableau 1 : Récupération de la mention OA via CrossRef.....   | 6         |
| Tableau 2 : Exemples de l'intégration de « esperluette ».....   | 7         |
| Tableau 3 : Exemples d'appauvrissement.....   | 7         |
| Tableau 4 : Exemples de traitements manuels.....  | 7         |

Figure 1 : Méthodologie du marquage des périodiques pour une année ..... 10  
Figure 2 : Méthodologie du marquage des notices pour une année..... 12

## Introduction

Afin de disposer d'éléments concrets pour éclairer la décision publique, le groupe de travail BSN4 a demandé à l'Inist la réalisation d'une étude bibliométrique sur les publications françaises en Open Access (OA) sur une période de 5 ans, de 2010 à 2014, afin d'en connaître le poids dans la production scientifique française et tenter d'estimer le montant des APCs.

Ce document présente l'approche méthodologique exhaustive mise en œuvre pour réaliser l'étude. Cette dernière est consultable :

- dans le document Word **OpenAccess\_BSN4\_EtudeBibliométrique\_30septembre2016.docx**.
- via le tableau de bord dynamique créé à l'aide de l'outil libre ezVIS<sup>1</sup>, dont l'accès est réservé.

Les conclusions et préconisations liées à la méthodologie sont à lire dans le document Word **OpenAccess\_BSN4\_Remarques\_Préconisations\_30septembre.docx**

Cette méthodologie met en évidence les différentes étapes qui ont permis de constituer le corpus de notices<sup>2</sup> à partir duquel les indicateurs de l'étude sont produits :

- constitution du Corpus France,
- marquage des périodiques,
- marquage des notices correspondant à des documents Open Access (OA),
- enrichissement des notices OA,
- constitution du Corpus France OA.

La dernière partie est consacrée aux indicateurs produits.

## Open Access

Le libre accès ou open access (OA) est la mise à disposition immédiate, gratuite et permanente sur internet des publications scientifiques issues de la recherche et de l'enseignement.

On distingue deux modèles ou voies de l'open access :

- la **voie verte** ou auto-archivage par les chercheurs des articles dans des archives électroniques ouvertes
- la **voie dorée (Gold Open Access)** ou publication d'articles dans une revue en libre accès, quel que soit son mode de financement. Une revue en libre accès répond aux exigences de qualité des articles par la présence d'un comité de lecture et possède un mode de financement permettant une diffusion aussi large que possible, sans restriction d'accès et d'utilisation.

On distingue trois types de revues en libre accès :

- des **revues entièrement gratuites** qui n'imposent pas de frais de publication ;
- des **revues entièrement Open Access** avec frais de publications « *article processing charge* » (APC) ;
- des **revues dites « Hybride »** accessibles sous abonnement et dans lesquelles un auteur est libre de publier des articles en Open Access, moyennant versement de frais de publication supplémentaires APC.

Dans les deux derniers cas, les frais de publications (APC) sont réglés en amont par l'auteur ou son institution. C'est un renversement du modèle économique traditionnel de l'édition avec passage d'un modèle lecteur-payeur où le lecteur paie par les abonnements... à un modèle auteur-payeur, où l'auteur, ou son institution, paie les coûts éditoriaux. Dans le cas des titres « Hybride », les institutions

---

<sup>1</sup> <https://github.com/madec-project/ezvis/>

<sup>2</sup> On emploiera indifféremment les termes de notices-publications-documents.

peuvent payer deux fois : d'une part les frais de publications pour la diffusion des articles en open access et d'autre part les abonnements aux revues.

Aussi aujourd'hui il est important de connaître l'impact de ce mode de diffusion scientifique et du modèle « Gold » sur les coûts de publications.

## Constitution du corpus France

La source utilisée pour constituer le Corpus France est le *Web of Science (WoS) Core Collection* de Thomson Reuters<sup>3</sup>. Pour les années de publication 2010 à 2014, les notices retenues :

- ont au moins un auteur affilié à un organisme situé en France (métropole et outre-mer)<sup>4</sup>,
- correspondent aux typologies de document suivantes : *articles, letters, proceedings papers* et *reviews*,
- appartiennent aux *Science Citation Index Expanded (SCI-EXPANDED)* et *Conference Proceedings Citation Index- Science (CPCI-S)*. Sont écartées les notices issues de *Social Sciences Citation Index (SSCI)*, de *Arts & Humanities Citation Index (A&HCI)*, de *Conference Proceedings Citation Index- Social Science & Humanities (CPCI-SSH)* et du *Book Citation Index*,
- ont au moins un ISSN. Les notices comportant uniquement un ISBN (sans ISSN) sont écartées.

## Repérage et marquage des périodiques OA ou « Hybride »

### Extraction et organisation des données

Pour chacune des 5 années, le travail suivant est réalisé.

A partir du corpus France, certains champs sont retenus pour chaque notice. Ils sont ensuite mis sous forme de tableau pour travailler à l'aide d'Omniscope<sup>5</sup> (Visokio Ltd, UK) :

- UT (*Unique Article Identifier*) : identifiant WoS, utile pour repérer un document.
- Titre du document.
- DOI (*Digital Object Identifier*) : utile pour le marquage via l'API (*Application Programming Interface*) CrossRef<sup>6</sup> qui permet de récupérer certaines mentions de licence.
- Année de publication : utile pour montrer l'évolution de la production OA.
- Titre du périodique : utile pour l'alignement avec les listes de périodiques de différentes sources. *Pour mémoire : un titre peut avoir plusieurs ISSN.*
- ISSN : utile pour l'alignement avec les listes de périodiques de différentes sources. *Pour mémoire : un ISSN peut être associé à plusieurs titres.*
- Titre de la série : utile pour l'alignement avec les listes de périodiques de différentes sources car c'est alors la série qui est considérée comme OA et non le titre spécifique d'un volume.
- Editeur : utile pour la répartition par éditeur.
- Affiliations : utile pour la répartition par organisme.
- Reprint Author (RP) : utile pour la répartition par organisme payeur des frais de publication.
- Type de document : utile pour la répartition par type de document.

<sup>3</sup> <http://thomsonreuters.com/en/products-services/scholarly-scientific-research/scholarly-search-and-discovery/web-of-science-core-collection.html>

<sup>4</sup> Equation de recherche : (CU=FRANCE OR CU=MARTINIQUE OR CU="FR POLYNESIA" OR CU="FRENCH GUIANA" OR CU=REUNION OR CU=GUADELOUPE OR CU="NEW CALEDONIA" OR CU=WALLIS OR CU=MAYOTTE)

<sup>5</sup> <http://www.visokio.com/>

<sup>6</sup> <http://www.crossref.org/>

## Répercussion du statut du périodique sur les publications

Pour le marquage OA des périodiques, suite à un premier test sur les années 2010 et 2014, il s'avère que les données de licence d'utilisation fournies pour une publication via un DOI par CrossRef (agence officielle d'enregistrement des DOIs) sont à ce jour incomplètes (voir Tableau 1). Il est donc décidé de **travailler au niveau des périodiques et de répercuter l'information sur tous les documents concernés**. Il s'agit de réduire au maximum la part de travail "manuel" à réaliser sur les sources susceptibles de fournir l'information au niveau des notices.

|   | OA   |      |
|---|------|------|
|   | 2010 | 2014 |
| Nb de publications considérées comme OA par le marquage des périodiques | 5042 | 8795 |
| Nb de publications avec DOI   | 3937 | 8197 |
| Nb de publications considérées comme OA par CrossRef                    | 868  | 2599 |
| Part des publications OA CrossRef / toutes publications confondues      | 17%  | 30%  |
| Part des publications OA CrossRef / publications avec DOI               | 22%  | 32%  |

Tableau 1 : Récupération de la mention OA via CrossRef

Les données pour les DOIs de publications issues de périodiques OA montrent bien l'incomplétude de CrossRef : en 2010, sur 5042 publications OA issues de périodiques OA, 3937 possèdent un DOI. A partir de ces DOIs, seules 868 publications sont associées à une mention OA via CrossRef.

### Traitement des périodiques

A partir du corpus France, les périodiques sont extraits.

Afin d'être le plus exhaustif possible sur le marquage OA, 5 sources d'information permettant de repérer les périodiques OA ou « Hybride » sont utilisées :

1. La liste des périodiques considérés comme OA dans le JCR (*Journal Citation Reports*<sup>7</sup> produit par Thomson Reuters) de chaque année<sup>8</sup>.
2. La liste de périodiques du DOAJ (*Directory of Open Access Journals*) au 30 mai 2016<sup>9</sup>. Elle précise l'année à partir de laquelle le périodique est devenu OA.
3. La liste des périodiques ROAD<sup>10</sup> (Répertoire des ressources scientifiques et universitaires en libre accès) en 2015.
4. La liste de périodiques OA de 7 éditeurs appelés dans l'étude « Grands Editeurs » (Elsevier, Springer, Taylor & Francis, Wiley Blackwell, Nature, IEEE, ACS) en 2015<sup>11</sup>.
5. La liste de périodiques « Hybride » des 7 « Grands Editeurs » ci-dessus en 2015.

Pour identifier les différents statuts des périodiques, des passes successives sont réalisées sur l'ensemble des périodiques en conservant tous les marquages attribués afin de détecter et d'évaluer d'éventuelles contradictions entre les différentes sources retenues. Les listes des périodiques du JCR et du DOAJ étant liées à une année, le marquage associé à ces listes est prioritaire. Il est à noter

<sup>7</sup> [http://wokinfo.com/products\\_tools/analytical/jcr/](http://wokinfo.com/products_tools/analytical/jcr/)

<sup>8</sup> Le résultat du filtre "Open Access" proposé par Thomson Reuters dans le WoS pour chaque notice n'a finalement pas été utilisé car la notion d'OA est attribuée à l'ensemble des publications du périodique quelle que soit l'année à laquelle le périodique est devenu OA : un périodique absent de la liste OA en 2010 et présent dans la liste OA 2011 voit ses publications 2010 considérées comme OA.

<sup>9</sup> <https://doaj.org/csv> : Cette liste est mise à jour toutes les 30 minutes.

<sup>10</sup> <http://road.ISSN.org/en>

<sup>11</sup> Les listes retenues de 2014 ou 2015 ne précisent pas le statut des périodiques pour les années antérieures.

cependant que si le périodique ne fait pas partie de ces 2 listes, la mention OA ou « Hybride » est affectée au périodique quelle que soit l'année de publication de 2010 à 2014.

## Traitement des titres des périodiques

Avant de comparer les titres des périodiques (ou les titres de série), des traitements tendant à les homogénéiser sont effectués.

Remplacement des and, y, und, et, & par esperluette

| Titre initial                              | Titre traité 1                                       |
|--|--|
| ENVIRONMENTAL SCIENCE & TECHNOLOGY LETTERS | ENVIRONMENTAL SCIENCE esperluette TECHNOLOGY LETTERS |
| JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY | JOURNAL OF AGRICULTURAL esperluette FOOD CHEMISTRY   |

Tableau 2 : Exemples de l'intégration de « esperluette »

Un programme, développé en interne, permet d'appauvrir les titres des périodiques :

- caractères minuscules
- caractères sans accent
- suppression de la ponctuation
- suppression des espaces
- suppression de caractères spéciaux (exemple : &)

| Titre traité 1   | Titre homogénéisé  |
|--|--|
| ENVIRONMENTAL SCIENCE esperluette TECHNOLOGY LETTERS                           | environmentalsciencesesperluette technologyletters                   |
| JOURNAL OF AGRICULTURAL esperluette FOOD CHEMISTRY                             | journalofagriculturalesperluettefoodchemistry                        |
| SEQUENCE-CONTROLLED POLYMERS: SYNTHESIS, SELF-ASSEMBLY, esperluette PROPERTIES | sequencecontrolledpolymerssynthesisselfassemblyesperluetteproperties |

Tableau 3 : Exemples d'appauvrissement

Des équivalences ont été repérées « visuellement » : sigle développé, article en début de titre ou absence d'article

| Titre initial WoS   | Titre équivalent d'une autre liste                      |
|---|---|
| BIOCHIMICA ET BIOPHYSICA ACTA-REVIEWS ON CANCER                 | Biochimica et biophysica acta (bba) - reviews on cancer |
| CALPHAD-COMPUTER COUPLING OF PHASE DIAGRAMS AND THERMOCHEMISTRY | Calphad   |
| AMERICAN JOURNAL OF CARDIOLOGY                                  | The american journal of cardiology                      |

Tableau 4 : Exemples de traitements manuels

On parlera par la suite de titres « homogénéisés ».

## Marquage des périodiques OA : JCR – DOAJ

Le marquage s'effectue par alignement des titres des périodiques « homogénéisés » et des ISSN du corpus initial WoS avec les titres et les ISSN des listes du JCR et du DOAJ de l'année concernée. Si les ISSN, ou les titres de périodiques, sont identiques, on attribue au périodique le marquage OA\_JCR et/ou OA\_DOAJ.

## Marquage des périodiques OA : ROAD – « Grands Editeurs »

Le marquage s'effectue par alignement des titres des périodiques « homogénéisés » et des ISSNs du corpus initial WoS avec les titres et les ISSNs des listes ROAD et des 7 « Grands Editeurs » de l'année 2015.

Si les ISSNs, ou les titres de périodiques, sont identiques, on attribue au périodique le marquage OA\_ROAD et/ou OA\_GrandsEditeurs.

## Marquage des périodiques « Hybride » des 7 « Grands Editeurs »

Le marquage s'effectue par alignement des ISSNs et des titres des périodiques « homogénéisés » « Hybride » des 7 « Grands Editeurs » avec les ISSNs et les titres du corpus.

Si les ISSNs, ou les titres de périodiques, sont identiques, on attribue au périodique le marquage Hybride\_GrandsEditeurs.

## Marquage des périodiques non OA des « Grands Editeurs »

Le marquage s'effectue par le repérage des périodiques qui n'ont été marqués ni OA ni « Hybride » lors des étapes précédentes et qui sont issus des 7 « Grands Editeurs ».

On attribue à ces périodiques le marquage NiNi\_GrandsEditeurs

## Marquage des périodiques sans statut

Le marquage s'effectue par le repérage des périodiques qui n'ont été marqués ni OA ni « Hybride » lors des étapes précédentes et qui ne sont pas issus des 7 « Grands Editeurs ».

On attribue à ces périodiques le marquage SansStatut.

## Résultat du marquage

Ces différentes passes réalisées à l'aide du logiciel Omniscope aboutissent au final à 4 types de « marquages » repérés par la suite par les codes couleur suivants :

- périodiques OA (OA\_JCR et/ou OA\_DOAJ ou OA\_ROAD et/ou OA\_GrandsEditeurs)
- périodiques « Hybride » des « Grands Editeurs » (Hybride\_GrandsEditeurs)
- périodiques ni OA, ni « Hybride » des « Grands Editeurs » (NiNi\_GrandsEditeurs)
- périodiques sans statut (SansStatut)

Le tableau final est constitué d'une ou plusieurs lignes par titre de périodique. Un titre peut avoir plusieurs ISSNs et un ISSN peut être associé à plusieurs titres.

| Issn      | Titre  |
|-----------|--|
| 0930-1038 | SURGICAL AND RADIOLOGIC ANATOMY  |
| 1279-8517 | SURGICAL AND RADIOLOGIC ANATOMY  |
| 1294-4475 | OIL & GAS SCIENCE AND TECHNOLOGY-REVUE D IFP ENERGIES NOUVELLES          |
| 1294-4475 | OIL & GAS SCIENCE AND TECHNOLOGY-REVUE DE L INSTITUT FRANCAIS DU PETROLE |

Table 1 : Exemples de titre avec plusieurs ISSN et d'ISSN avec plusieurs titres

A ce niveau, un contrôle est effectué pour vérifier qu'un titre de périodique a toujours le même marquage, et qu'un ISSN a toujours le même marquage. En effet, le couple « titre du périodique1 / ISSN électronique » peut être marqué OA alors que le couple « titre du périodique1 / ISSN papier » peut être considéré comme non OA. Cette situation se présente quand l'alignement s'effectue seulement au niveau d'un ISSN. De même, le couple « titre du périodique1 / ISSN1 » peut être marqué OA alors que le couple « titre du périodique2/ ISSN1 » peut être considéré comme non OA. Cette situation se présente quand l'alignement s'effectue au niveau d'un titre seulement. Les ajustements nécessaires sont réalisés.



La méthodologie du marquage des périodiques est résumée dans le schéma suivant.

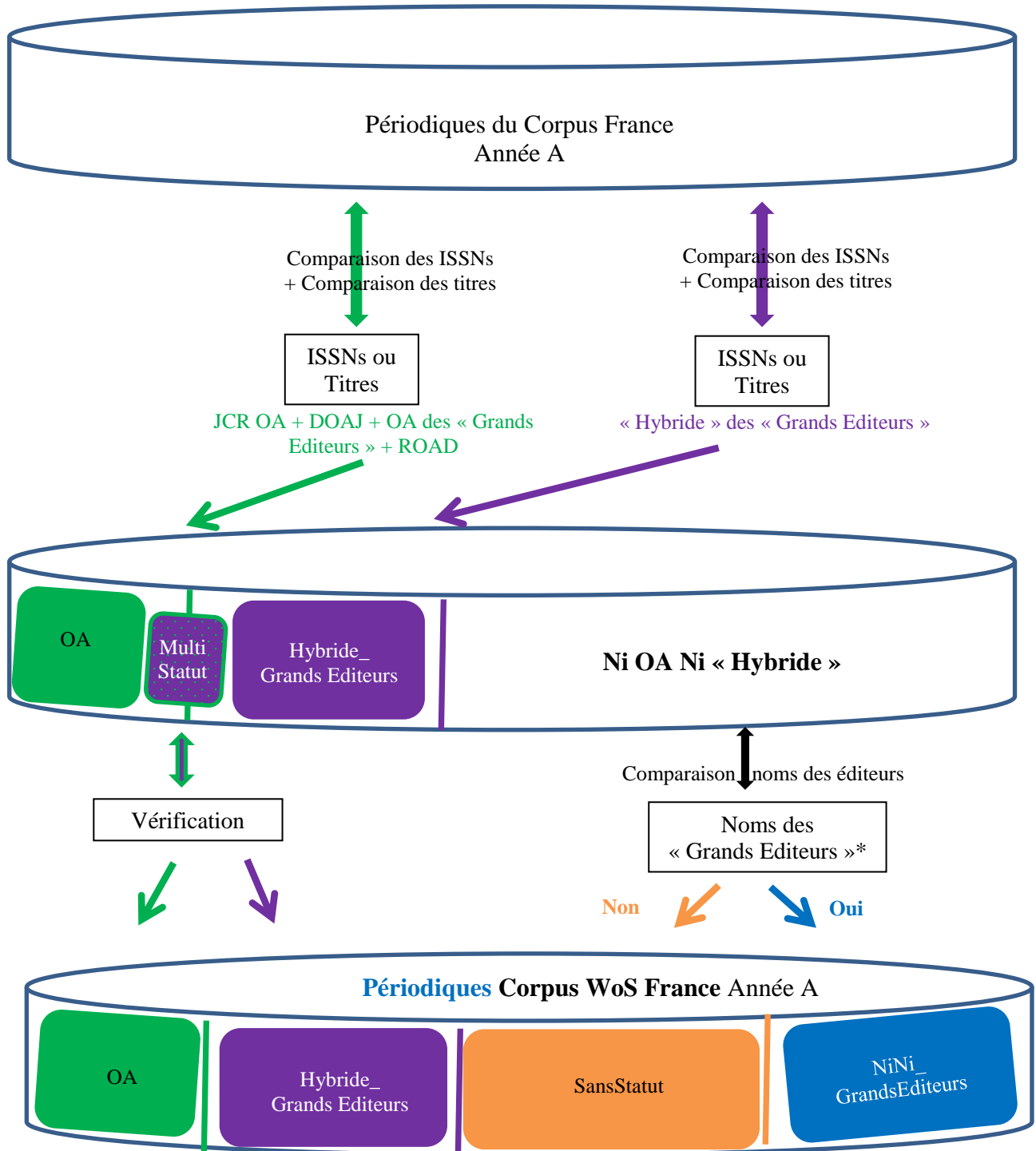


Figure 1 : Méthodologie du marquage des périodiques pour une année

\* Elsevier, Springer, Taylor & Francis, Wiley Blackwell, Nature, IEEE, ACS : si l'éditeur du périodique fait partie de cette liste, le périodique est marqué « NiNi\_GrandsEditeurs » (pour Ni OA Ni « Hybride ») ; si l'éditeur du périodique ne fait pas partie de cette liste, le périodique est marqué « SansStatut ».

## Marquage des notices OA

### Répercussion du marquage des périodiques OA

Le marquage OA effectué au niveau des périodiques est répercuté au niveau des notices. L'alignement se fait à partir du couple « titre du périodique-ISSN » présent dans la notice et celui présent dans le tableau des périodiques final.

### Traitement des notices issues de périodiques « Hybride » ou de périodiques « SansStatut »

Si les parts « OA » et « NiNi\_GrandsEditeurs » ne nécessitent a priori pas de traitement supplémentaire, il en va autrement pour les parts « Hybride\_GrandsEditeurs » et « SansStatut ». Pour ces deux parts, un traitement par le DOI est indispensable pour récupérer l'information via CrossRef. Pour consolider la part « NiNi\_GrandsEditeurs », ses DOIs sont également traités.

Les DOIs présents dans les notices sont isolés pour interroger l'API CrossRef et récupérer les mentions de licence. Les deux valeurs « Doi-Licence-Content-Version » et « doi-license-URL » retournées par l'API sont exploitées. Les notices dont le DOI est associé à la valeur VoR (*Version of Record*) de la première valeur et aux urls de licence suivantes de la seconde sont considérées comme OA :

- <http://creativecommons.org/license/by/3.0/>
- <http://creativecommons.org/licenses/by/2.5/za/>
- <http://creativecommons.org/licenses/by/3.0/>
- [http://creativecommons.org/licenses/by/3.0/deed.en\\_US](http://creativecommons.org/licenses/by/3.0/deed.en_US)
- <http://creativecommons.org/licenses/by/4.0/>
- <http://creativecommons.org/licenses/by-nc/3.0/>
- <http://creativecommons.org/publicdomain/zero/1.0/>
- [http://pubs.acs.org/page/policy/authorchoice\\_ccby\\_termsofuse.html](http://pubs.acs.org/page/policy/authorchoice_ccby_termsofuse.html)
- [http://pubs.acs.org/page/policy/authorchoice\\_cbyncnd\\_termsofuse.html](http://pubs.acs.org/page/policy/authorchoice_cbyncnd_termsofuse.html)
- [http://pubs.acs.org/page/policy/authorchoice\\_termsofuse.html](http://pubs.acs.org/page/policy/authorchoice_termsofuse.html)
- <https://creativecommons.org/publicdomain/zero/1.0/>

Les notices issues de périodiques « OA » sont alors marquées OA\_OA.

Les notices issues de périodiques « Hybride\_GrandsEditeurs » et repérées comme étant OA via CrossRef sont alors marquées OA\_Hybride.

Les notices issues de périodiques « NiNi\_GrandsEditeurs » ou « SansStatut » et repérées comme étant OA via CrossRef sont alors marquées OA\_Indéterminés. Les périodiques sont vérifiés manuellement et les notices sont ensuite marquées OA\_OA ou OA\_Hybride.

Le schéma suivant résume le marquage des publications du Corpus France.

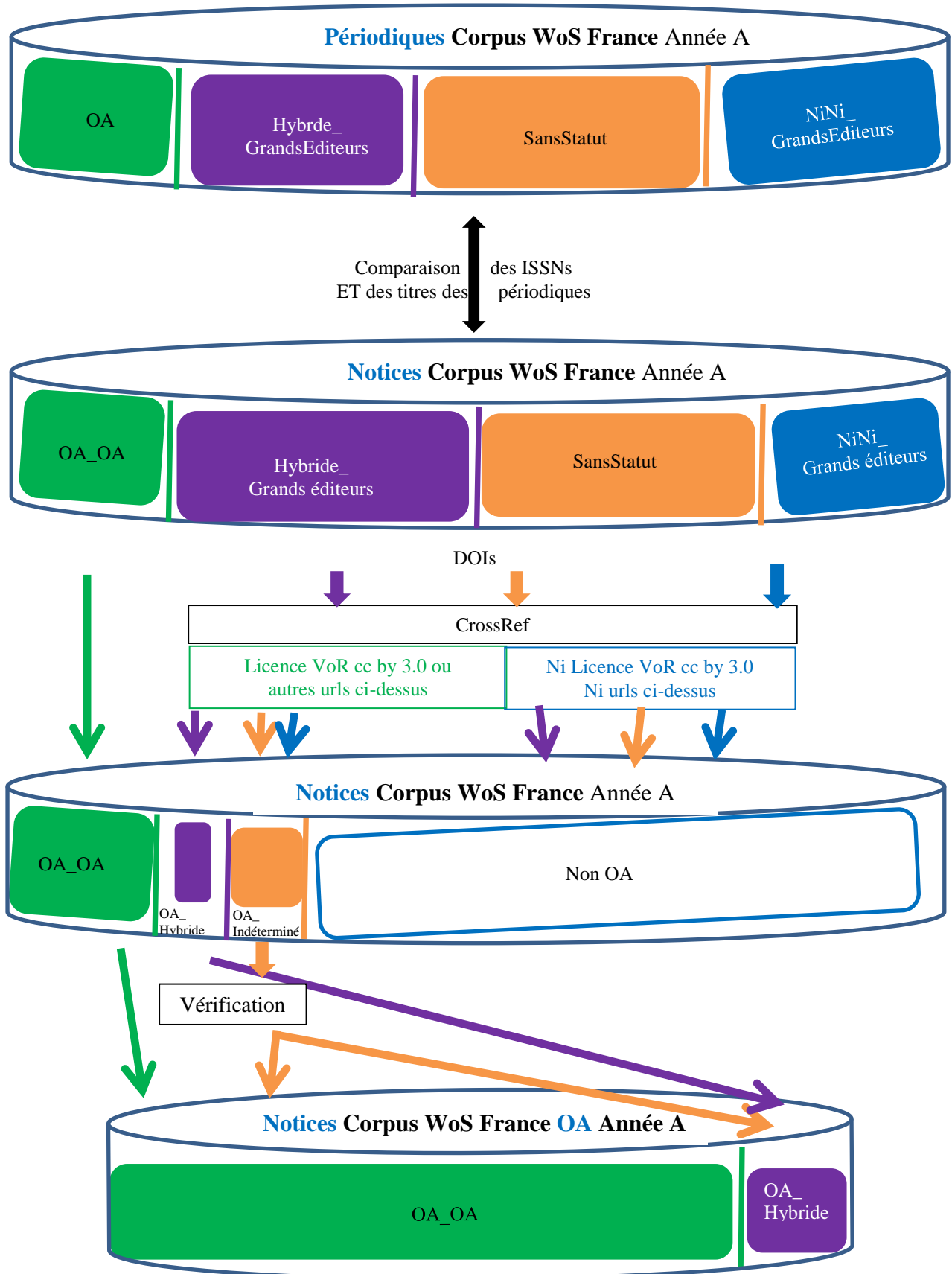


Figure 2 : Méthodologie du marquage des notices pour une année  
 (La taille des blocs donne une idée des volumes mais ne représente pas la répartition proportionnelle réelle).

## Enrichissement des notices OA

Un enrichissement des notices est indispensable avant de constituer le corpus définitif à partir duquel l'étude bibliométrique sera réalisée.

### Marquage thématique

Titres et ISSNs des notices OA (OA\_OA et OA\_Hybride) sont alignés avec titres et ISSNs du JCR de l'année concernée pour récupérer les « *subject categories* »<sup>12</sup>. A partir de ces dernières, des équivalences sont établies avec les 22 disciplines ESI<sup>13</sup>.

Il est utile de rappeler que :

- tous les périodiques n'ont pas de « *subject category* » (titres absents du JCR),
- une « *subject category* » est attribuée à un périodique et non à une publication,
- un périodique peut se voir attribuer plusieurs « *subject categories* »,
- un périodique peut changer de « *subject category* » d'une année sur l'autre.

### Homogénéisation des titres des périodiques

Comme il en a été question plus haut, un même ISSN peut être associé à plusieurs titres de périodique. L'étude bibliométrique présentant la répartition des publications par titre, il est indispensable qu'une homogénéisation soit effectuée. Une table d'équivalence est donc créée.

| Titre initial   | Titre final  |
|---|--|
| OIL & GAS SCIENCE AND TECHNOLOGY-REVUE DE L<br>INSTITUT FRANCAIS DU PETROLE | Oil & Gas Science And Technology-Revue D<br>Ifp Energies Nouvelles |
| OIL & GAS SCIENCE AND TECHNOLOGY-REVUE D IFP<br>ENERGIES NOUVELLES          | Oil & Gas Science And Technology-Revue D<br>Ifp Energies Nouvelles |

Table 2 : Extrait d'une table d'équivalence de titres de périodiques

### Homogénéisation des éditeurs

L'écriture du nom des éditeurs peut différer d'une notice à l'autre. L'étude bibliométrique présentant la répartition des publications par éditeur, il est indispensable qu'une homogénéisation soit effectuée. Une table d'équivalence est donc créée.

| Editeur initial                     | Editeur final        |
|-------------------------------------|----------------------|
| CZECH ACAD AGRIC SCI                | Czech Acad Agric Sci |
| CZECH ACADEMY AGRICULTURAL SCIENCES | Czech Acad Agric Sci |

Table 3 : Extrait d'une table d'équivalence d'éditeurs

<sup>12</sup> Les « *subjects categories* » sont des thématiques attribuées pour chaque périodique présent dans le JCR.

<sup>13</sup> Les 22 disciplines ESI sont créées par Thomson Reuters dans le cadre d'*Incites Essential Science Indicators* : elles sont constituées à partir de regroupement de « *subject categories* ».

## Attribution des APCs

Pour calculer le coût de l'OA, les données APC (*Article Processing Charge*) sont récupérées à l'aide de la liste du DOAJ et d'une liste de périodiques fournie par l'Inserm.

### DOAJ

La liste de périodiques du DOAJ propose 2 données dédiées aux APC : l'une avec le montant de l'APC, l'autre avec la monnaie. A partir de ces informations, le montant de l'APC est calculé en € et en \$ en fonction du taux de change du moment.

Il est nécessaire de signaler que le montant de l'APC n'est pas associé à une année particulière. Il est donc identique pour les 5 années traitées.

### Fichier Inserm

Pour compléter les informations du DOAJ (près des trois quarts des périodiques du corpus France OA n'ont pas de donnée APC), l'Inserm a fourni ses données en € et en \$. Celles-ci ont été générées dans un fichier Excel à partir des données éditeurs, disponibles soit sous forme de listes avec le montant des APCs pour chaque revue soit sur les sites de chaque revue d'un éditeur.

Pour ces données, comme pour celles du DOAJ, le montant de l'APC n'est pas associé à une année particulière.

### Traitement des APCs

Quand un titre de périodique a un APC présent dans les 2 listes, le montant de la liste proposée par l'Inserm est retenu.

Pour information, il peut y avoir des écarts conséquents entre les montants des 2 listes.

### Traitement des champs « Affiliations » et « Reprint Author »

Un point de la demande initiale du groupe BSN4 est le calcul de la part des publications OA et de leur coût par organisme. Pour y répondre, il est envisagé de récupérer le marquage détenu par l'OST (Observatoire des Sciences et Techniques). Or ce dernier n'est pas en mesure de le fournir.

Des traitements, au départ non prévus, sont donc réalisés à partir des champs « Affiliations » et « Reprint Author ».

### Champ « Affiliations »

Seules les affiliations françaises (métropole et outre-mer) sont conservées. Sont recherchées, dans ces affiliations, les chaînes de caractères correspondant à 5 organismes choisis a priori :

- CEA
- CNRS
- Inra
- Inserm
- ParisTech

Quand les chaînes de caractères sont trouvées, la notice est enrichie du marquage du nom de l'organisme correspondant. Une même notice peut se voir attribuer plusieurs marquages si plusieurs organismes de la liste ci-dessus sont concernés.

## Champ « Reprint Author »

Ce champ est exploité car il est considéré comme élément permettant de connaître l'organisme payeur des frais de publication.

Les notices avec un « *Reprint Author* » (RP) français (métropole et outre-mer) sont marquées, dans un nouveau champ « Pays RP », France, les autres Non France. Un autre champ est créé « APC RP » : il servira à calculer le coût de l'OA. Il contient le montant APC uniquement pour les notices « Pays RP = France ». Les notices où « Pays RP = Non France » n'ont pas de montant dans cette colonne.

Dans ce champ « *Reprint Author* », sont également recherchées, les chaînes de caractères correspondant aux 5 organismes choisis a priori ci-dessus. Quand les chaînes de caractères sont trouvées, la notice est enrichie du marquage du nom de l'organisme RP correspondant. Une même notice peut se voir attribuer plusieurs marquages si plusieurs organismes RP de la liste ci-dessus sont concernés.

## Vérification des données périodiques

A ce niveau, un nouveau contrôle est effectué pour vérifier qu'un titre de périodique a toujours les mêmes marquages (« *subject categories* », disciplines ESI, APC). En effet, le couple « titre du périodique1 / ISSN électronique » peut avoir un montant alors que le couple « titre du périodique1 / ISSN papier » peut en avoir un autre. Les ajustements nécessaires sont effectués.

## Constitution du corpus OA final

Une fois l'ensemble des précédents traitements réalisé pour chacune des 5 années, les 5 corpus annuels sont fusionnés.

De nouvelles homogénéisations et vérifications sont alors effectuées.

## Homogénéisation des thématiques

Si un périodique a différentes « *subject categories* » en fonction de l'année, cela ne pose pas de problème et rien n'est modifié. Par contre si un périodique possède des « *subject categories* » une année et n'en possède pas une autre année, l'information connue est répercutée sur toutes les années.

## Homogénéisation des titres

De la même manière que les titres ont été homogénéisés au sein d'une année, ils le sont lors de la concaténation des 5 années.

## Homogénéisation des éditeurs

De la même manière que les éditeurs ont été homogénéisés au sein d'une année, ils le sont lors de la concaténation des 5 années.

## Homogénéisation des données APC

La vérification des données s'effectue au niveau du montant des APCs qui doit être identique pour toutes les années SAUF si un périodique est devenu OA pendant la période quinquennale. La date de publication est alors comparée à l'année à partir de laquelle le périodique est devenu OA selon le DOAJ.

## Format de la notice finale d'une publication

Une notice finale est constituée des éléments initiaux et des **éléments ajoutés** suite aux traitements de marquage des périodiques et d'enrichissement des notices sous forme de tableau :

- Titre de la publication
- UT
- Année de publication
- **Statut de la publication**
- Titre du périodique - Titre de la série
- Titre du périodique initial (correspond au titre du volume du périodique faisant partie d'une série)
- Editeur
- ISSN
- Type de document
- **Subject Categories**
- **Disciplines ESI**
- **Montant de l'APC €** qui sert à calculer le coût de l'OA.
- **Montant de l'APC DOAJ \$**
- **Montant de l'APC Inserm \$**
- Reprint Author
- **Pays RP**
- **Organisme RP**
- **Affiliations françaises**
- **Statut du périodique**
- **Année OA** (année à partir de laquelle le périodique est devenu OA selon le DOAJ)
- **URL de la licence fournie par CrossRef** : avec, sous ezVIS, la possibilité d'accéder à l'adresse
- DOI : avec, sous ezVIS, la possibilité d'accéder au document

## Statistiques

Une fois le corpus final prêt, les données sont traitées :

- dans Excel pour calculer les indicateurs et créer les graphiques nécessaires,
- dans ezVIS pour créer les tableaux de bord dynamiques  
<http://bsn4-oasept2016.board.inist.fr/index.html>  
login utilisateur : bsn4  
mot de passe : bsn4\_openaccess

Ces outils permettent de mettre en forme les données pour répondre, pour chacune des années et pour la période 2010-2014, aux points suivants :

- Nombre et part des publications françaises OA par rapport à la production scientifique française
- Répartition des publications françaises OA entre les publications issues de périodiques OA et les publications issues de périodiques « Hybride » (statut de publication)
- Répartition des publications françaises OA par type de document
- Répartition des publications françaises OA par éditeur en distinguant les publications issues de périodiques OA de celles issues de périodiques « Hybride »
- Répartition des publications françaises OA par périodique en distinguant les publications issues de périodiques OA de celles issues de périodiques « Hybride »
- Répartition des publications françaises OA par thématique
  - par « *subject category* » : suite aux premiers résultats fournis, il n'a pas été jugé utile d'intégrer cette répartition dans l'Etude bibliométrique.



- par disciplines ESI
- Coût des publications françaises OA pour la France
- Coût et répartition des publications françaises OA pour les 5 organismes retenus (CEA, CNRS, Inra, Inserm, ParisTech)
- Coût des publications françaises OA par éditeur en distinguant le coût des publications issues de périodiques OA de celles issues de périodiques « Hybride »
- Coût des publications françaises OA par périodique en distinguant le coût des publications issues de périodiques OA de celles issues de périodiques « Hybride ».