



## Rapport d'étape sur l'analyse des instructions émises par les établissements à l'attention de leurs chercheurs concernant les publications scientifiques

Odile Hologne, Laure Martineau, Suzy Ramanana

### ► To cite this version:

Odile Hologne, Laure Martineau, Suzy Ramanana. Rapport d'étape sur l'analyse des instructions émises par les établissements à l'attention de leurs chercheurs concernant les publications scientifiques. [0] 2007, 23 p. hal-02824665

**HAL Id: hal-02824665**

**<https://hal.inrae.fr/hal-02824665>**

Submitted on 6 Jun 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Rapport d'étape sur l'analyse des instructions émises par les établissements à l'attention de leurs chercheurs concernant les publications scientifiques

---

*Projet NormAdresses – OST, Groupe de travail « Analyse des instructions »*

### Membres du groupe :

Nicole Haeffner-Cavaillon (INSERM), Dominique Cavet (IRD), Agnès Raymond Denise (Institut Pasteur), Sandrine Mouret (Collège de France), Christine Boucquiaux (AMUE), Jean-Michel Torrenti (LCPC), Laure Martineau (CPU), Annie Buffeteau-Hejblum (IFP), Patricia Volland (INRA), Marie-Christine Lambert (CIRAD), Jacques Millet (INRIA), Lucile Grasset (MRT)

**Animatrice** : Odile Hologne (INRA)

**Rédacteurs** : Odile Hologne (INRA), Laure Martineau (CPU), Suzy Ramanana (OST)

### *Avertissement au lecteur : positionnement du présent document, suites à donner*

*Ce document résulte des travaux du groupe de travail « Analyse des instructions faites aux chercheurs ». Ce groupe a étudié les instructions émises par 16 établissements en direction de leurs chercheurs afin d'en dresser une typologie. Il a ensuite analysé l'impact potentiel de ces instructions sur la retranscription et l'interprétation des adresses dans le Web of Science.*

*Le groupe formule des propositions pour améliorer les pratiques actuelles et explicite les questions qui nécessitent des arbitrages qui ne sont pas de son ressort, notamment parce qu'elles concernent les règles à appliquer pour le libellé des adresses des laboratoires mixtes.*

*Le contexte du projet est détaillé dans le document « GT\_Transversal\_Normadresses.pdf » accessible sur le site de l'OST.*

### **Contexte du projet**

*Depuis 2005, dans un contexte marqué par la mise en œuvre de la LOLF et la diffusion internationale de plusieurs classements des universités et organismes de recherche, tel que le « classement de Shanghai », les établissements de recherche français sont de plus en plus soucieux de la bonne identification de leur production scientifique.*

## **Enjeux**

*De nombreuses institutions de recherche ont donc donné des instructions à leurs chercheurs et enseignants chercheurs pour retranscrire leur affiliation dans les publications. Ces instructions, qui visent à ordonner les éléments d'une adresse et à rendre visibles les structures de recherche, sont parfois contradictoires entre les tutelles d'une même UMR. La nature de leur impact sur les outils automatisés d'analyse bibliométrique, notamment ceux construits à partir des données du Web of Science (WoS) de la société Thomson Scientific, doit être analysée, comprise, puis diffusée à tous les décideurs et acteurs du système de recherche français. En effet, le WoS fait référence mondialement pour la production d'indicateurs bibliométriques. C'est une des sources de données du « classement de Shanghai ». Son éditeur, Thomson Scientific, commercialise des outils d'analyse bibliométrique automatisée tel que Essential Science Indicators (ESI) pour valoriser ces données<sup>1</sup>. Enfin, la base de données de Thomson ISI est l'outil de travail de l'OST pour la production de tous ses indicateurs bibliométriques.*

---

<sup>1</sup> les articles du journal Les Echos de janvier et mars 2007, évoquant le classement mondial des organismes de recherche français, font référence à des données issues de ESI

## Sommaire

1. Définitions.....	4
2. La retranscription des affiliations dans le <i>Web of Science</i> .....	5
3. Elaboration d'une typologie des instructions faites par les différentes institutions .	7
4. Analyse des conséquences potentielles de ces choix dans le WOS et dans ESI....	8
4.1 Le sens d'écriture des adresses dépend souvent des revues scientifiques ou des pratiques des communautés disciplinaires .....	8
4.2 Conséquences potentielles du choix mono-ligne ou multi-ligne .....	9
<i>Modèle multi-ligne</i> .....	9
<i>Modèle mono-ligne</i> .....	11
<i>Multi-ligne / mono-ligne : des conséquences différentes pour les UMR</i> .....	12
5. Analyse des conséquences potentielles pour le calcul des indicateurs .....	14
5.1 Quelques rappels sur les indicateurs .....	14
5.2 Incidence du mode d'écriture des affiliations sur le calcul des indicateurs.....	15
<i>Incidence sur les indicateurs au niveau de la visibilité des institutions</i> .....	15
<i>Incidence sur les indicateurs LOLF</i> :.....	18
<i>Incidence sur les indicateurs nationaux</i> .....	19
6. Propositions du groupe de travail .....	22
6.1 Les 5 propositions techniques consensuelles .....	22
6.2 Les points particuliers à considérer pour maximiser la visibilité dans ESI .....	23

## 1. Définitions

### Affiliation et adresse

Des liens explicites entre les auteurs et leurs affiliations sont toujours présents dans les publications : un système d'indices liés à chaque auteur renvoie sur un ensemble d'adresses, représentant les affiliations.

#### Affiliation :

L'affiliation d'un auteur décrit son appartenance à une structure de recherche : l'unité et les tutelles de l'unité. Le nombre de partenaires peut varier de  $\{1 \dots n\}$  : 1 dans le cas d'une unité propre, n dans le cas d'une unité ayant plusieurs tutelles (organismes, universités, etc.).

#### Adresse :

C'est la forme utilisée dans les publications et dans les bases de données pour représenter les affiliations.

#### Hiérarchie ascendante (écriture d'une adresse selon une) :

C'est le mode d'écriture d'une adresse réalisé en commençant par le nom du laboratoire.  
*Nom du laboratoire, organisme, code postal, ville, pays*

#### Hiérarchie descendante (écriture d'une adresse selon une) :

C'est le mode d'écriture d'une adresse en commençant par le nom de l'organisme de recherche.

*Organisme, nom du laboratoire, code postal, ville, pays*

#### Affiliation mono-ligne ou multi-ligne :

Les auteurs peuvent indiquer leur affiliation à plusieurs tutelles en utilisant des indices qui permettent de faire des renvois aux différentes adresses :

(i) Un seul renvoi par auteur pour décrire l'appartenance à une unité ayant « n » tutelles : on parle alors de l'affiliation « mono-ligne » : chaque auteur renvoie à une seule adresse (autrement dit, chaque affiliation est représentée par une seule adresse).

Dupont B.<sup>1</sup>, Durand D.<sup>1</sup>, Martin L.<sup>2</sup>, Smith J.L.<sup>3</sup>

<sup>1</sup> Université de Limoges, CNRS, UMR 6172, F-87000 Limoges, France

<sup>2</sup> INRA, CIRAD, UMR 1026 Pathologie, F-34000 Montpellier, France

<sup>3</sup> Massachusset Institut of Technologie, ....., USA

(ii) « n » renvois par auteur dans le cas de « n » tutelles : on parle alors du « multi-lignes » : chaque auteur renvoie à plusieurs adresses (autrement dit, chaque affiliation est représentée par plusieurs adresses).

Ce modèle permet de décrire explicitement chacune des tutelles des UMR dans une adresse séparée.

Dupont B.<sup>1,2</sup>, Durand D.<sup>1,2</sup>, Martin L.<sup>3,4</sup>, Smith J.L.<sup>5</sup>

<sup>1</sup> CNRS, UMR 6172, F-87000 Limoges, France

<sup>2</sup> Université de Limoges, UMR 6172, F-87000 Limoges, France

<sup>3</sup> INRA, UMR 1026 Pathologie, F-34000 Montpellier, France

<sup>4</sup> CIRAD, UMR 1026 Pathologie, F-34000 Montpellier, France

<sup>5</sup> Massachusset Institut of Technologie, ....., USA

## 2. La retranscription des affiliations dans le *Web of Science*

La mise en base des adresses chez *Thomson Scientific* reste encore relativement artisanale, elle est en partie au moins réalisée par des indexeurs à partir des scans des articles des revues fournis par les éditeurs.

*Thomson Scientific* structure les adresses dans ses bases de données selon un modèle hiérarchique descendant [main organization] [sub organization] ... [country]. Il peut y avoir plusieurs [sub organization] mais une seule [main organization].

position adresses	adresse	organisation	sous org 1	sous org2	sous org3
1	GANIL, DSM, CEA, IN2P3,CNRS, F- 14021 Caen, France	GANIL	DSM	CEA	IN2P3,CNRS
2	SPHN Saclay, DAPNIA, DSM, CEA, F-91191 Gif Sur Yvette, France	SPHN Saclay	DAPNIA	DSM	CEA
3	Univ Ioannina, Dept Phys, GR- 45110 Ioannina, Greece	Univ Ioannina	Dept Phys		
4	Univ Sao Paulo, Inst Fis, DFN, BR-01498 Sao Paulo, Brazil	Univ Sao Paulo	Inst Fis	DFN	

Les étapes de traitement des adresses chez *Thomson* sont décrites ci-dessous :

1. Numérisation des revues et traitement par des indexeurs
2. « Normalisation » pour structuration canonique : recherche de l'organisation principale et hiérarchie descendante

Dans le WoS, les liens entre les auteurs et les adresses ne sont pas mémorisés ; les travaux de bibliométrie utilisant le WoS se fondent donc uniquement sur les **adresses**, chaque publication pouvant disposer de une ou plusieurs adresses.

Le modèle canonique est calqué sur l'organisation de la recherche américaine :

*Main Organization (Main Org) (Organisation principale)*

*Definition: The Main Organization is defined as the largest unit of authority in any given address. It will be processed in the Main Org field.*

*Sub-organization (Sub Org)*

*Definition: A subdivision is any section of the address other than the main organization, street information, city, state, country and postal code or zip. It can include institutions that are organized within a parent, or main institution, departments, divisions, named laboratories, etc.*

**Thomson Scientific met quasi systématiquement en « Organisation principale » l'Université, quelle que soit la place qu'elle occupe dans l'adresse. Il n'y a qu'une seule "organisation principale" par adresse. Il n'y a aucun lien entre un auteur et une adresse dans le WOS (sauf pour l'auteur de correspondance : reprint author).**

Les traitements bibliométriques automatiques des outils Essential science indicators (ESI) ou la fonction « Analyze » du Wos exploitent les données du champ [main organization].

Addresses: CNRS, UNIV MONTPELLIER 2, CHIM BIOORGAN LAB, SCI & TECH LANGUEDOC, URA488, F-34095 MONTPELLIER 5, France

<input type="button" value="VIEW RECORDS"/>	<b>Field: Institution Name</b>	<b>Record Count</b>	<b>% of 1</b>	<b>Bar Chart</b>	<input type="button" value="SAVE ANALYSIS DATA TO FILE"/>
<input type="checkbox"/>	CNRS	1	100.0000 %	<div style="width: 100%; height: 10px; background-color: blue;"></div>	
<input type="button" value="VIEW RECORDS"/>					<input type="button" value="SAVE ANALYSIS DATA TO FILE"/>

### 3. Elaboration d'une typologie des instructions faites par les différentes institutions

Les instructions émises par 16 institutions de recherche (organismes ou universités) ont mis en évidence des points de convergence, et un critère permettant de différencier ces instructions selon 2 critères :

- L'ordre d'écriture des structures d'appartenance : s'agit-il d'une hiérarchie descendante ou ascendante ?
- La façon de structurer l'adresse des unités mixtes : selon le modèle mono-ligne ou multi-ligne ?

Typologie	Mono-ligne	Multi-ligne
Ascendant	CNRS	
Descendant	INP Lorraine Université Nancy 1 Université Marne la Vallée Inria IRD	Inserm Cemagref Cirad Université Paris Sud Université Clermont 1 Université Metz Institut Cochin Université Pierre et Marie Curie Université Paris 5 Université Montpellier

La typologie permet de dégager 3 grands types d'instructions données par les organismes :

- Ascendant/mono-ligne : 1 cas
- Descendant /mono-ligne : 5 cas
- Descendant / multi-ligne : 10 cas



## 4. Analyse des conséquences potentielles de ces choix dans le WOS et dans ESI

### 4.1 Le sens d'écriture des adresses dépend souvent des revues scientifiques ou des pratiques des communautés disciplinaires

Les journaux, dans leurs notes aux auteurs, peuvent proposer un modèle d'écriture des affiliations. Certaines communautés disciplinaires ont également leurs pratiques. Ainsi, en physique, c'est souvent le laboratoire qui est mis en premier.

Lors de la retranscription d'une affiliation d'un auteur, les équipes d'indexeur du WoS identifient la « main organization », si l'adresse comporte d'autres structures ou organismes, ils sont vus comme des [sub organization]

Par exemple, l'affiliation d'un auteur d'une unité de recherche :

Martin, L.<sup>1</sup>

- <sup>1.</sup> Unité de physiologie animale, INRA, F-35000 Rennes

ou

- <sup>1.</sup> INRA, Unité de physiologie animale, F-35000 Rennes

sera retraduite de façon identique dans le WOS selon un modèle hiérarchique descendant :

INRA, Unité de physiologie animale, F-35000 Rennes, France.

[main organization]=INRA

[sub organization]=Unité de physiologie animale

[city]=Rennes

[zip code]=35000

[country]=France

**Remarque :** Thomson Scientific peut avoir du mal à repérer la [main organization]. Ainsi dans l'exemple ci-dessous, qui mélange écriture ascendante et descendante, le CNRS n'est jamais vu comme une [main organization] alors que le laboratoire est identifié comme tel.

Author(s): Valton, S. Peyrin, F. Delpierre, P. Debarbieux, F. Bonissent, A. Khoury, R. Meessen, C. Morel, C. Sappey-Marinier, D.

Addresses: Sappey-Marinier D (reprint author), CREATIS, CNRS, UMR 5515, F-69621 Villeurbanne, France

**CREATIS**, CNRS, UMR 5515, F-69621 Villeurbanne, France





**CREATIS**, INSERM, U630, F-69621 Villeurbanne, France

**Inst Natl Sci Appl**, **CREATIS**, F-69621 Villeurbanne, France

Univ Lyon 1, **CREATIS**, F-69677 Bron, France

Univ Mediterranee Aix Marseille 2, CPPM, CNRS, INP3, UMR 6550, F-13288 Marseille, France

### *Lancement de la fonction « Analyze » sur cette publication*

<input type="checkbox"/> VIEW RECORDS	Field: Institution Name	Record Count	% of 1	Bar Chart
<input type="checkbox"/>	CREATIS	1	100.0000 %	
<input type="checkbox"/>	INST NATL SCI APPL	1	100.0000 %	
<input type="checkbox"/>	UNIV LYON 1	1	100.0000 %	
<input type="checkbox"/>	UNIV MEDITERRANEE AIX MARSEILLE 2	1	100.0000 %	
<input type="checkbox"/> VIEW RECORDS				

## 4.2 Conséquences potentielles du choix mono-ligne ou multi-ligne

Nous avons approfondi l'analyse de plusieurs exemples pour comprendre les conséquences d'une écriture mono ou multi-ligne en partant des articles puis en étudiant l'enregistrement bibliographique dans le WOS et sa répercussion sur la fonction de comptage « Analyze » du WoS (qui fonctionne selon les mêmes principe que ESI) . Les cas présentés ci-dessous sont représentatifs des 2 modèles multi ou mono-ligne.

### Modèle multi-ligne

#### *Extrait d'un article où les auteurs ont appliqué le modèle multi-ligne*

Les affiliations sont précisées avec des renvois sous la forme d'indices <sup>1,2,3</sup> (la présentation sous forme condensée comme ci-dessous ou avec une ligne par indice n'a aucune incidence sur le résultat).

**Yohann Duverger<sup>1,2,3</sup>, Jérôme Belougne<sup>4</sup>, Sarah Scaglione<sup>1,2,3</sup>,  
Dominique Brandli<sup>4</sup>, Christophe Beclin<sup>4</sup> and Jonathan J. Ewbank<sup>1,2,3,\*</sup>**

<sup>1</sup> Centre d'Immunologie de Marseille-Luminy, Université de la Méditerranée  
Case 906, 13288 Marseille cedex 9, France <sup>2</sup> INSERM U631, 13288  
Marseille, France <sup>3</sup> CNRS UMR6102, 13288 Marseille, France <sup>4</sup> CNRS,  
Institut de Biologie du Développement de Marseille-Luminy Marseille, France

\* : auteur de correspondance (reprint author)

Y. Duverger, S. Scaglione et J.J. Ewbank ont indiqué leur appartenance à une UMR à trois tutelles (Université de la Méditerranée, INSERM, CNRS) avec 3 renvois par auteur.




### Enregistrement correspondant dans le WOS

Chaque établissement (CNRS, INSERM, Université) est identifié comme [main organisation], le centre d'immunologie comme une [sub organization] de l'université

**Addresses:** Ewbank JJ (reprint author), Univ Mediterranee, Ctr Immunol, Case 906, F-13288 Marseille 9, France  
Univ Mediterranee, Ctr Immunol, F-13288 Marseille 9, France  
**INSERM**, U631, F-13288 Marseille, France  
CNRS, UMR6102, F-13288 Marseille, France  
CNRS, Inst Dev Biol, Marseille, France

### Lancement de la fonction « Analyze » sur cette publication

Cette fonction affecte les publications aux établissements identifiés comme [main organization]. Il s'agit d'un compte de présence.

<input type="checkbox"/>	Field: Institution Name	Record Count	% of 1	Bar Chart	<input type="button" value="VIEW RECORDS"/>	<input type="button" value="SAVE ANALYSIS DATA TO FILE"/>
<input type="checkbox"/>	CNRS	1	100.0000 %			
<input type="checkbox"/>	INSERM	1	100.0000 %			
<input type="checkbox"/>	UNIV MEDITERRANEE	1	100.0000 %			
<input type="button" value="VIEW RECORDS"/>						<input type="button" value="SAVE ANALYSIS DATA TO FILE"/>

NB : Nous avons pu observer que des affiliations libellées sous la forme suivante : Dupond, A.<sup>1</sup>

1. UMR Bioclimatologie, Université Bordeaux 1 AND Bordeaux 2, F-33000 Bordeaux

généraient 2 lignes d'adresses dans le WOS et pouvaient donc être considérées comme équivalentes au modèle à renvois multiples.

### Modèle mono-ligne

*Extrait d'un article où les auteurs ont appliqué le modèle mono-ligne*

**T. Cadoudal<sup>1</sup>, J. M. Blouin<sup>1</sup>, M. Collinet<sup>1</sup>, F. Fouque<sup>1</sup>, G. D. Tan<sup>2</sup>, E. Loizon<sup>3</sup>, E. G. Beale<sup>4</sup>, K. N. Frayn<sup>2</sup>, F. Karpe<sup>2</sup>, H. Vidal<sup>3</sup>, C. Benelli<sup>1</sup> and C. Forest<sup>1</sup>✉**

(1) INSERM UMR-S 530/747; Université Paris Descartes, Centre Universitaire des Saints-Pères, 45 rue des Saints-Pères, F-75006 Paris, France

(2) Oxford Centre for Diabetes Endocrinology and Metabolism, Churchill Hospital, University of Oxford, Oxford, UK

(3) INSERM, U449; INRA 1235; Université Claude Bernard, Faculté de Médecine R. Laennec, Lyon, France

(4) Department of Cell Biology and Biochemistry, Texas Tech University Health Sciences Center, Lubbock, TX, USA

### *Enregistrement correspondant dans le WOS*

Thomson a réordonné les adresses au profit des universités. L'université Claude Bernard a été retranscrite sous l'appellation « Univ Lyon 1 ».

**Addresses:** Forest C (reprint author), Univ Paris Descartes, INSERM, UMR S 530 747, Ctr Univ St Peres, 45 Rue St Peres, F-75006 Paris, France


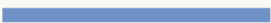


Univ Paris Descartes, INSERM, UMR S 530 747, Ctr Univ St Peres, F-75006 Paris, France

Univ Oxford, Churchill Hosp, Oxford Ctr Diabet Endocrinol & Metab, Oxford, England

Univ Lyon 1, Fac Med R Laennec, **INSERM**, U449, **INRA** 1235, F-69365 Lyon, France

Texas Tech Univ, Hlth Sci Ctr, Dept Biochem & Cell Biol, Lubbock, TX 79409 USA

### *Lancement de la fonction « Analyze » sur cette publication*

<input type="button" value="VIEW RECORDS"/>	Field: Institution Name	Record Count	% of 1	Bar Chart	<input type="button" value="SAVE ANALYSIS DATA TO FILE"/>
<input type="checkbox"/>	TEXAS TECH UNIV	1	100.0000 %		
<input type="checkbox"/>	UNIV LYON 1	1	100.0000 %		
<input type="checkbox"/>	UNIV OXFORD	1	100.0000 %		
<input type="checkbox"/>	UNIV PARIS DESCARTES	1	100.0000 %		
<input type="button" value="VIEW RECORDS"/>					<input type="button" value="SAVE ANALYSIS DATA TO FILE"/>

Dans un produit comme ESI ou dans la fonction « Analyze » du WOS, cette publication ne sera jamais attribuée à l'INSERM ou à l'INRA bien que ces organismes soient dans les tutelles des UMR car ils sont identifiés comme des [sub organization]. Par contre, pour les indicateurs de la LOLF, le repérage fait par les organismes et les universités, dans le cadre de l'OST à partir de la base complète de *Thomson Scientific*, permet d'attribuer ces publications aux bons opérateurs grâce à un travail minutieux de repérage des adresses par chacun.

### Multi-ligne / mono-ligne : des conséquences différentes pour les UMR

Pour les UMR, les deux systèmes d'écriture ont des impacts différents en matière de visibilité dans des produits comme ESI.

	Mono-ligne	Multi-ligne
Avantages	<ul style="list-style-type: none"><li>• La notion d'UMR est sauvegardée</li><li>• La distinction entre tutelle et coopération entre organismes est possible</li><li>• Conformité avec les pratiques habituelles de l'OST</li></ul>	<ul style="list-style-type: none"><li>• Toutes les tutelles d'une UMR sont correctement identifiées</li><li>• Toutes les tutelles sont visibles dans ESI</li></ul>
Inconvénients	<ul style="list-style-type: none"><li>• Seule une tutelle (en général l'université) est prise en compte dans ESI</li></ul>	<ul style="list-style-type: none"><li>• La notion d'UMR est perdue.</li><li>• Il est impossible de distinguer tutelle d'une unité et collaboration entre ces tutelles</li><li>• Non conformité avec les pratiques habituelles de l'OST</li></ul>

Le modèle multi-ligne présente l'avantage d'être facilement lisible par *Thomson Scientific*. On peut observer dans le WOS que les établissements ainsi cités sont bien retranscrits dans la base ce qui leur permet de se voir attribuer la publication dans le produit ESI ou dans la fonction « Analyze » du WOS.

Le tableau ci-dessous montre la « perte de visibilité » des organismes français. Cette perte de visibilité peut se traduire à partir de la dernière colonne : pour l'INSERM et le CNRS de l'ordre de 80%, pour l'INRA de l'ordre de 30%

La requête faite dans le WOS a permis d'interroger à la fois sur les champs [main organization] et [sub organization] (colonne 2), alors que dans ESI (colonne 3) seules les données du champ [main organization] sont traitées par l'application.

## Publications référencées dans le *Web of Science* entre 1996 et 2006 (hors sciences sociales)

Requête du WOS: (PY=1996-2006) and (AD=institution), résultats dans la colonne N° 2

Organisme	Publications repérées dans le WoS	Publications retenues dans la base ESI	Pourcentage retenu dans ESI
CNRS	201 860	48 701	24%
INSERM	67 313	11 117	17%
CEA	29 113	23 876	82%
INRA	26 697	19 014	71%
Inst. Pasteur France	13 901	11 827	85%
IRD	6 370	2 669	42%
MNHN	4 890	3 632	74%
Institut Curie	4 886	3 334	68%
INRIA	4 061	3 726	92%
Collège de France	3 480	2 838	82%
IFREMER	2 840	2 196	77%
CIRAD	2 353	1 669	71%

Source : Service en ligne *Web of Knowledge* (SCIE et ESI) consulté le 15 mars 2007 – calculs IRD.

- Pour le WoS, les calculs ont été effectués par l'IRD.
- Pour l'IRD, il s'agit des publications avec des adresses IRD ou Orstom.

*Commentaire : Dans l'hypothèse où nous demanderions à Thomson de ne plus interpréter /traduire les adresses françaises selon ses règles, le modèle mono-ligne oblige à choisir la [main organization] et à utiliser un modèle hiérarchique descendant pour positionner la [main organization] en premier. Le modèle multi-ligne ne comporte pas ces contraintes.*

Le modèle multi-ligne pose des problèmes pour la pertinence du calcul de certains indicateurs, notamment les indicateurs en compte fractionnaire et les indicateurs de collaboration. En effet, il devient impossible avec les données du WOS, de distinguer les adresses des institutions tutelles d'UMR et les adresses des institutions qui collaborent dans le cadre de projets de recherche. Nous illustrons ces problèmes ci-dessous.

## 5. Analyse des conséquences potentielles pour le calcul des indicateurs

Dans cette partie, nous analysons les conséquences des différentes pratiques d'écriture des affiliations sur la valeur des indicateurs en compte fractionnaire ou en compte de présence calculés par l'OST. Nous montrons également l'impact sur la visibilité des organismes dans le produit ESI.

### 5.1 Quelques rappels sur les indicateurs

Tout calcul d'indicateur est dépendant à la fois du niveau d'unification (ou du niveau d'exhaustivité d'identification des acteurs) et du choix du type de compte.

Il existe plusieurs niveaux d'unification. Certains sont réalisés par *Thomson ISI* ou l'OST et touchent à la structuration de la base : ils conditionnent la comparabilité des indicateurs entre les acteurs identifiés. D'autres sont dépendants de l'intégration dans la base *Thomson ISI* (indexation manuelle actuellement) et du recensement par les acteurs de leur propre production : ils conditionnent l'exhaustivité des résultats d'interrogation.

En particulier la reconnaissance de la production par une institution pose le problème du périmètre de reconnaissance en particulier pour les UMR. Nous y reviendrons.

L'OST et d'autres producteurs d'indicateurs utilisent deux types de comptes pour calculer les indicateurs.

#### Compte de présence

En compte de présence, les indicateurs répondent aux deux questions suivantes :

- Combien de publications peut-on attendre de la participation d'un acteur ?
- Combien de publications peut-on attendre de la collaboration entre deux acteurs ?

On identifie en premier l'article que l'on attribue pour part entière à chacune des institutions identifiées dans l'affiliation.

Dans ce cas, l'interprétation est immédiate mais les comptes ne peuvent être consolidés. L'indicateur s'interprète en termes de **participation**, la place de chaque acteur dans la production globale est mesurée indépendamment de celle des autres.

#### Compte fractionnaire

La question à laquelle répondent les indicateurs calculés en compte fractionnaire est la suivante :

- Au niveau d'un article publié, quelle est la part de contribution de chaque signataire ?

Un article est alors considéré comme une unité. On attribue à chaque institution une part de cet article. D'interprétation moins immédiate, ce compte est consolidable à tous niveaux. Les indicateurs s'interprètent en termes de **contribution**, sans double compte entre acteurs. La place de chaque acteur dans la production commune est mesurée en relation avec celle des autres.

De manière concrète, à partir de l'article suivant identifié :

Dupont B.<sup>1,2</sup>, Durand D.<sup>1,2</sup>, Martin L.<sup>3,4</sup>, Smith J.L.<sup>5</sup>

<sup>1</sup> CNRS, UMR 6172, F-87000 Limoges, France

<sup>2</sup> Université de Limoges, UMR 6172, F-87000 Limoges, France

<sup>3</sup> INRA, UMR 1026 Pathologie, F-34000 Montpellier, France

<sup>4</sup> CIRAD, UMR 1026 Pathologie, F-34000 Montpellier, France

<sup>5</sup> Massachusetts Institut of Technologie, ....., USA

	Nombre d'articles publiés par l'institution	
Institutions	Calculé en compte de présence	Calculé en compte fractionnaire
CNRS	1	1/5
Université de Limoges	1	1/5
INRA	1	1/5
CIRAD	1	1/5
MIT	1	1/5

NB : le calcul des indicateurs porte sur les adresses et pas sur les individus (dans le WOS il est impossible de faire le lien entre un auteur et son affiliation).

## 5.2 Incidence du mode d'écriture des affiliations sur le calcul des indicateurs

### Incidence sur les indicateurs au niveau de la visibilité des institutions

A partir d'un article signé par quatre auteurs payés chacun par des institutions différentes et appartenant à des laboratoires communs ou une institution étrangère suivant ce schéma :

Dupont B. : UMR 6172, F-87000 Limoges, France, chercheur CNRS

Durand D. : UMR 6172, F-87000 Limoges, France, enseignant chercheur de l'université de Limoges

Martin L. : UMR 1026 Pathologie, F-34000 Montpellier, France, chercheur INRA mais dans une UMR INRA/CIRAD

Smith J.L. : Massachusset Institut of Technologie, ....., USA



Nous avons identifié (au moins) quatre modalités différentes pour indiquer l'affiliation. Le modèle de base reste le mono-ligne ou le multi-ligne mais les consignes politiques données aux auteurs peuvent générer des variantes :

#### **Cas n° 1 : Multi-ligne**

Dupont B.<sup>1,2</sup>, Durand D.<sup>1,2</sup>, Martin L.<sup>3,4</sup>, Smith J.L.<sup>5</sup>

<sup>1</sup> CNRS, UMR 6172 (ou UMR6172), F-87000 Limoges, France

<sup>2</sup> Université de Limoges, UMR 6172, F-87000 Limoges, France

<sup>3</sup> INRA, UMR 1026 Pathologie, F-34000 Montpellier, France

<sup>4</sup> CIRAD, UMR 1026 Pathologie, F-34000 Montpellier, France

<sup>5</sup> Massachusetts Institut of Technologie, ....., USA

#### **Cas n° 2 : Mono-ligne**

Dupont B.<sup>1</sup>, Durand D.<sup>1</sup>, Martin L.<sup>2</sup>, Smith J.L.<sup>3</sup>

<sup>1</sup> Université de Limoges, CNRS, UMR 6172, F-87000 Limoges, France

<sup>2</sup> INRA, CIRAD, UMR 1026 Pathologie, F-34000 Montpellier, France

<sup>3</sup> Massachusetts Institut of Technologie, ....., USA

#### **Cas n° 3 : Mono-ligne avec consigne de mentionner en premier l'institution d'appartenance du signataire :**

Dupont B.<sup>1</sup>, Durand D.<sup>2</sup>, Martin L.<sup>3</sup>, Smith J.L.<sup>4</sup>

<sup>1</sup> CNRS, Université de Limoges, UMR 6172, F-87000 Limoges, France

<sup>2</sup> Université de Limoges, CNRS, UMR 6172, F-87000 Limoges, France

<sup>3</sup> INRA, CIRAD, UMR 1026 Pathologie, F-34000 Montpellier, France

<sup>4</sup> Massachusetts Institut of Technologie, ....., USA

#### **Cas n° 4 : Seule l'institution d'appartenance du signataire est indiquée :**

Dupont B.<sup>1</sup>, Durand D.<sup>2</sup>, Martin L.<sup>3</sup>, Smith J.L.<sup>4</sup>

<sup>1</sup> CNRS, UMR 6172, F-87000 Limoges, France

<sup>2</sup> Université de Limoges, UMR 6172, F-87000 Limoges, France

<sup>3</sup> INRA, UMR 1026 Pathologie, F-34000 Montpellier, France

<sup>4</sup> Massachusetts Institut of Technologie, ....., USA

Les cas 3 et 4 correspondent à des conséquences différentes pour la revendication des publications.



Attention, ces calculs reposent sur des hypothèses purement théoriques.

Ils sont conçus pour le calcul des tableaux de bord institutionnels : chaque institution est prise en compte pour elle-même.

Nous pouvons recalculer nos indicateurs en compte de présence et en compte fractionnaire, et analyser la prise en compte par ESI. Lors des calculs d'indicateurs pour une institution donnée, l'OST regroupe l'ensemble des partenaires de cette institution en une seule entité « XXX\_COLLAB ». Nous obtenons, pour chaque institution participant à l'article, l'ensemble de tableaux suivants :

	Compte de présence								Compte fractionnaire			
	Cas 1		Cas 2		Cas 3		Cas 4		Cas 1	Cas 2	Cas 3	Cas 4
	OST	ESI	OST	ESI	OST	ESI	OST	ESI	OST	OST	OST	OST
CNRS	1	1	1	0	1	1	1	1	1/5	1/3	1/2	1/4
CNRS_COLLAB	1		1		1		1		4/5	2/3	1/2	3/4

	Compte de présence								Compte fractionnaire			
	Cas 1		Cas 2		Cas 3		Cas 4		Cas 1	Cas 2	Cas 3	Cas 4
	OST	ESI	OST	ESI	OST	ESI	OST	ESI	OST	OST	OST	OST
Université. Limoges	1	1	1	1	1	1	1	1	1/5	1/3	1/2	1/4
Université. Limoges_COLLAB	1		1		1		1		4/5	2/3	1/2	3/4

	Compte de présence								Compte fractionnaire			
	Cas 1		Cas 2		Cas 3		Cas 4		Cas 1	Cas 2	Cas 3	Cas 4
	OST	ESI	OST	ESI	OST	ESI	OST	ESI	OST	OST	OST	OST
INRA	1	1	1	1	1	1	1	1	1/5	1/3	1/4	1/4
INRA_COLLAB	1		1		1		0		4/5	2/3	3/4	3/4

	Compte de présence								Compte fractionnaire			
	Cas 1		Cas 2		Cas 3		Cas 4		Cas 1	Cas 2	Cas 3	Cas 4
	OST	ESI	OST	ESI	OST	ESI	OST	ESI	OST	OST	OST	OST
CIRAD	1	1	1	0	1	0	0	0	1/5	1/3	1/4	0
CIRAD_COLLAB	1		1		1		1		4/5	2/3	3/4	0

	Compte de présence								Compte fractionnaire			
	Cas 1		Cas 2		Cas 3		Cas 4		Cas 1	Cas 2	Cas 3	Cas 4
	OST	ESI	OST	ESI	OST	ESI	OST	ESI	OST	OST	OST	OST
MIT	1	1	1	1	1	1	1	1	1/5	1/3	1/4	1/4
MIT_COLLAB	1		1		1		1		4/5	2/3	3/4	3/4

Nous voyons ici que suivant le mode d'écriture retenu, certaines institutions ne seront pas visibles dans ESI. De plus, nous commençons à mesurer la complexité d'interprétation de ces indicateurs en compte fractionnaire dont la valeur dépend des modalités d'écriture des affiliations retenues par les institutions.

Le tableau montre une incidence en compte de présence uniquement dans le cas n°4, une des institutions n'apparaissant plus.

#### **Incidence sur les indicateurs LOLF :**

Le CNRS émerge à l'un des programmes de la MIREs (P1), tandis que l'université émerge à un autre (P2), le CIRAD et l'INRA à un troisième (P3), tandis que le MIT n'est pas concerné par la LOLF (ext).

Lors d'une production d'indicateurs, chaque programme est pris en compte pour lui-même, c'est à dire que les opérateurs de ce programme sont regroupés en une seule entité (« méta-établissement »).

Recalculons nos indicateurs en compte fractionnaire au niveau des programmes. On a alors l'ensemble de tableaux suivants :

	Indicateurs LOLF			
	Cas 1	Cas 2	Cas 3	Cas 4
P1	1/5	1/3	2/4	1/4
P1_COLLAB	4/5	2/3	2/4	3/4

	Indicateurs LOLF			
	Cas 1	Cas 2	Cas 3	Cas 4
P2	1/5	1/3	2/4	1/4
P2_COLLAB	4/5	2/3	2/4	3/4

	Indicateurs LOLF			
	Cas 1	Cas 2	Cas 3	Cas 4
P3	2/5	1/3	1/4	1/4
P3_COLLAB	3/5	2/3	3/4	3/4

Là encore il peut y avoir une grande incidence sur les valeurs des indicateurs.

Les différentes consignes données aux chercheurs et enseignants chercheurs en matière d'affiliation introduisent dès maintenant des biais dans les valeurs des indicateurs. Nous serons dans l'incapacité d'interpréter les variations pluriannuelles de ces valeurs : la baisse des comptes fractionnaires peut être due à une diminution de la part de l'acteur ou à l'impact du modèle multi-ligne...

Si nous nous positionnons au niveau « Pays » pour calculer ces indicateurs, nous observons cette fois un effet inflationniste sur le compte fractionnaire puisque dans le cas 2 (mono-ligne), la part France est de 2/3 alors que dans le cas 1 (multi-ligne) elle passe à 4/5.

L'OST a effectué des simulations pour démontrer cet impact.

#### *Incidence sur les indicateurs nationaux*

L'OST a calculé les indicateurs à partir des données de sa base en dupliquant des lignes pour simuler une instruction uniforme d'affiliation de type multi-ligne (cas 1). Cette approche purement théorique met en évidence l'effet inflationniste par rapport au mono-ligne (cas 2) pour les indicateurs de production ou de visibilité.

Ceci est illustré dans les tableaux suivants :

### Indicateurs de production - Parts mondiales de publications de la France

		ANNEE			
Discipline	Type de compte	2001	2002	2003	2004
Toutes disciplines confondues					
	Présence	6,43	6,33	6,15	5,97
	Fractionnaire	4,99	4,86	4,67	4,46
	Simulation fractionnaire avec doublon d'adresses françaises en SDV	5,15	5,02	4,82	4,62
SDV					
	Présence	5,97	5,76	5,61	5,44
	Fractionnaire	4,91	4,67	4,49	4,30
	Simulation fractionnaire avec doublon d'adresses françaises en SDV	5,19	4,96	4,78	4,59

### Indicateurs de visibilité - Impacts relatifs de la France par rapport au monde

		ANNEE			
Discipline	Type de compte	2001	2002	2003	2004
Toutes disciplines confondues					
	Fractionnaire	0,93	0,91	0,92	0,95
	Simulation fractionnaire avec doublon d'adresses françaises en SDV	0,96	0,95	0,96	0,98
SDV					
	Fractionnaire	0,90	0,91	0,90	0,93
	Simulation fractionnaire avec doublon d'adresses françaises en SDV	0,95	0,95	0,94	0,97

## Indicateurs de partenariat - Profil de collaboration en compte de présence

Toutes disciplines confondues (2002)

	Publications totales (a)	Publications sans partenariat (b)	Publications en partenariat (c)	Publications en partenariat France-France (d)	Publications en partenariat international (e)
Nombre	48967	15266	33701	19363	20126
Pct	100,0	31,2	68,8	39,5	41,1
Impact à 2 ans*	1,00	0,63	1,17	1,04	1,41

Simulation avec doublon d'adresses françaises en SDV					
Nombre	48967	7610	41357	32514	20126
Pct	100,0	15,5	84,5	66,4	41,1
Impact à 2 ans*	1,00	0,56	1,08	1,09	1,41

(a)=(b)+(c)

(d) et (e) ne sont pas additifs à cause des partenariats multiples

(\*) ratio entre profil de citations et profil de publications

*Ces tableaux montrent les conséquences des modes d'écriture des affiliations sur les valeurs des indicateurs. Implicitement ils posent la question de la robustesse du système d'information des indicateurs car leur qualité et donc leur sens dépendront de plus en plus des traitements qui seront effectués ou non sur les adresses.*

## 6. Propositions du groupe de travail

Cinq propositions techniques consensuelles peuvent être formulées.

### 6.1 Les 5 propositions techniques consensuelles

**RT1 : les opérateurs de recherche et en particulier les universités doivent choisir une appellation unique et à communiquer auprès des chercheurs pour qu'ils utilisent uniquement cette forme « normalisée ».**

*Commentaire : Pour les universités, l'existence de multiples formes (Université Paris 5, Université René Descartes ...) pour une même université ou entre universités nuit à la lisibilité des opérateurs à l'international. Le numéro des universités peut être perçu comme un classement. Nous préconisons une harmonisation des appellations.*

**RT2 : L'auteur principal doit s'efforcer d'harmoniser le modèle d'écriture des affiliations et opter uniformément pour une hiérarchie ascendante ou descendante pour l'ensemble de ses co-auteurs.**

**RT3 : Les adresses doivent être simples et courtes. Il n'est pas nécessaire qu'elles soient toutes retranscrites dans leur intégralité, ni qu'elles comportent les cedex – il est recommandé d'identifier un site principal et utiliser le code postal de la ville correspondante (sans cedex). Seul l'auteur principal indiquerait son adresse postale complète.**

**RT4 : le libellé des adresses des publications des UMR pose de nombreuses questions. En effet, beaucoup de choix seraient théoriquement possibles, et les évaluer n'est pas du ressort du présent groupe de travail. Cependant, il est clair qu'il serait très souhaitable de définir des règles communes claires.**

**RT5 : Une affiliation ne peut pas refléter toutes les dimensions du pilotage ou des financements de la recherche de l'opérateur, elle ne devrait mentionner que des structures ayant une pérennité et une stabilité dans le temps.**

*Commentaire : Autrement dit, le soutien d'un projet par l'ANR, un pôle de compétitivité, l'Europe, etc., ne devrait pas apparaître dans l'affiliation, mais dans la rubrique appropriée de la notice.*

## 6.2 Les points particuliers à considérer pour maximiser la visibilité dans ESI

Des points particuliers doivent être pris en considération si l'on souhaite maximiser la visibilité des institutions dans des produits d'analyse bibliométrique automatique tel que ESI de *Thomson Scientific*.

**RT6 : Il faudrait donner à l'ISI une liste des [main organization] françaises, gérer sa mise à jour et indiquer dans une table de correspondance le lien entre dénomination officielle et variantes du nom. On peut envisager cette mise à disposition des données via le futur référentiel des structures de recherche françaises.**

*Commentaire : De la même manière, si les PRES décident d'apparaître comme une institution unique, ils devront être listés en [main organisation], et les universités ou écoles qui les constituent devront être notées en [sub organization]- ou ne plus être mentionnées dans les affiliations (à arbitrer).*

**RT7 : Le produit ESI semble poser les mêmes problèmes aux structures de recherche de nombreux pays, en Europe ou en Asie. Une autre piste pourrait être d'inciter *Thomson Scientific* à faire évoluer son modèle de données pour prendre en compte des structures de recherche plus complexes ayant plusieurs [main organization].**

**RT8 : Si *Thomson Scientific* ne change pas son modèle de données pour prendre en compte plusieurs [main organization], le modèle d'affiliation « multi-ligne » est celui qui garantit la meilleure visibilité de tous les opérateurs dans ESI. Il présente l'inconvénient de fausser les valeurs des indicateurs de la LOLF (indicateurs en compte fractionnaire\* et indicateurs de collaboration). Un choix de priorités est à faire.**

(\*) Type de compte utilisé aussi par d'autres producteurs d'indicateurs comme la NSF :

National Science Board. 2006. *Science and Engineering Indicators 2006*, Volume1 (5-37), Volume 2 (Appendix Table 5-41, 5-42) [<http://www.nsf.gov/statistics/seind06/pdf/volume1.pdf>, <http://www.nsf.gov/statistics/seind06/pdf/volume2.pdf>]

Ces dernières remarques montrent les choix à faire si l'on veut maximiser, au niveau des institutions, la visibilité des opérateurs de recherche français dans les classements internationaux institutionnels. Elles prennent en compte le fait que ces classements utilisent souvent les résultats fournis automatiquement par *Thomson Scientific*. Les conséquences de leur mise en œuvre, et notamment le choix des modèles multi-ligne ou mono-ligne, doivent être évalués par rapport à cet objectif très spécifique.

Nous suggérons une mise en œuvre rapide des propositions techniques consensuelles avec une stratégie de diffusion nationale adaptée à l'enjeu.