



Module 2 Séquence 2

Le nommage des fichiers

Fred de Lamotte - Montpellier
<https://orcid.org/0000-0003-4234-1172>

Julien Seiler - Strasbourg
@julzi



Le nommage des fichiers



COMMENT

Open Access



Gene name errors are widespread in the scientific literature

Mark Ziemann¹, Yotam Eren^{1,2} and Assam El-Osta^{1,3*}

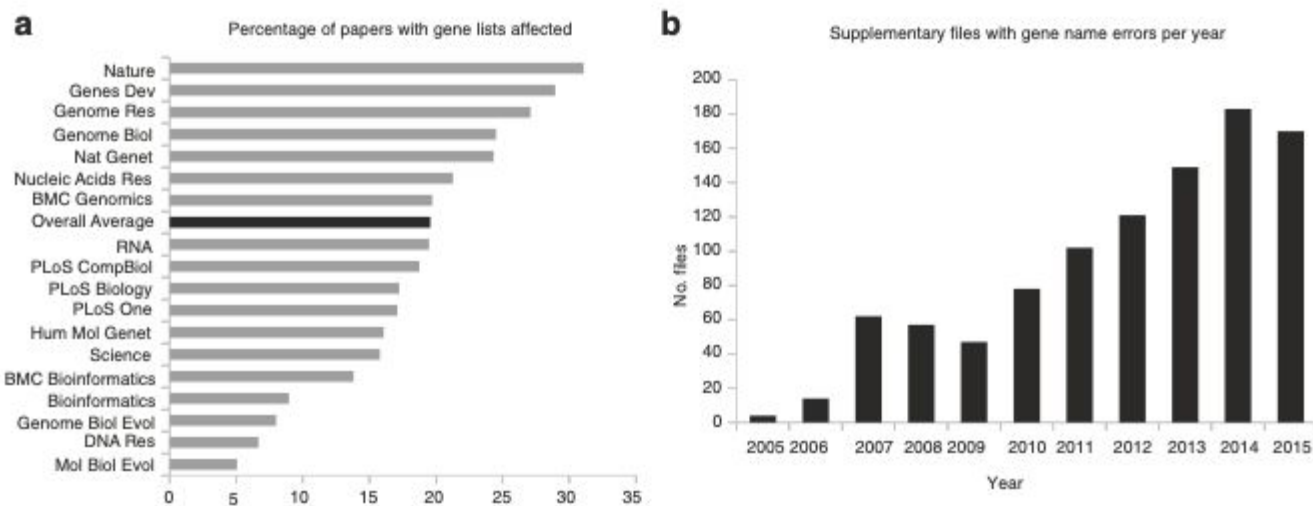


Fig. 1 Prevalence of gene name errors in supplementary Excel files. **a** Percentage of published papers with supplementary gene lists in Excel files affected by gene name errors. **b** Increase in gene name errors by year

1) DONNER UN NOM BREF ET EXPLICITE (SI VOUS N'ÊTES PAS UNE MACHINE)

Le nom d'un fichier doit être le plus court possible et doit refléter son contenu. Il peut contenir son **sujet** (rapport d'activité, résultats d'analyse, bilan, etc.), sa **date**, sa **version** (si le document est destiné à évoluer).

Ce qui signifie qu'il est inutile d'écrire des mots vides comme : "le", "un", "des", "et", etc. Ils n'apportent pas plus de sens pour comprendre de quoi le document retourne.

Pour réduire le nombre de caractères, utilisez des **abréviations** facilement reconnaissables (CR pour compte rendu par exemple).



Reunion



20150407_CR



20150407_Minutes



20150407_OJ

Dans le dossier réunion, on retrouve :

un document de compte rendu (CR)
un document sur les minutes de la réunion et
un document sur l'ordre du jour (OJ).

Le mot réunion n'apparaît pas dans la dénomination des fichiers car il est déjà précisé dans le nom du dossier.

2) NE PAS METTRE D'ESPACE NI DE CARACTÈRES SPÉCIAUX (*SURTOUT SI VOUS ÊTES UNE MACHINE*)

Il vaut mieux ne pas mettre d'espace entre les mots ni de caractères spéciaux (accents et cédille y compris). Pour séparer les mots, vous pouvez mettre une **majuscule** ou un **underscore** “_”.



Règles dénomination fichiers



ReglesDenominationFichiers



Cette règle permet d'éviter des problèmes d'affichage voire d'accès aux documents.

En effet, il se peut que d'un système d'exploitation à un autre (Windows vers Mac par exemple), le fichier ne soit pas reconnu.

Il vaut mieux donc bannir tout ce qui n'est pas alphanumérique : & / + > : ? % (] etc.

3) INDIQUER LES DATES AU BON FORMAT (*QUE VOUS SOYEZ OU NON UNE MACHINE*)

Les dates doivent s'écrire au **format AAAAMMJJ** (année, mois, jour).

Commencer par l'année permet de **maintenir les documents dans un ordre chronologique**.



20150405_CR



20160310_CR



20160515_CR

Le document le plus récent se retrouvera toujours en fin de liste.

4) PLACER L'ÉLÉMENT IMPORTANT EN PREMIER

Il est préférable de bien choisir l'ordre des éléments dans le nom des fichiers. Mettez **en premier l'élément qui vous est le plus pertinent** pour pouvoir retrouver facilement votre document.



20150911_CR



20150911_Minutes



20151210_CR



20151210_Minutes

Ici par exemple, il sera plus facile de retrouver un compte rendu en fonction de sa date. L'élément le plus important du document est sa date, c'est donc elle qui apparaît en premier.

5) INDIQUER LES VERSIONS DES DOCUMENTS

Certains documents peuvent être amenés à évoluer, avoir plusieurs versions avant une version finale. Dans ces cas, il faut préciser la version dans le nom du fichier.



Convention_V01



Convention_V02



Convention_VF

Pour faire simple, ne rédigez que la lettre V suivi du numéro de la version, en deux unités.

Si un numéro de version ne convient pas, vous pouvez le remplacer par une autre convention de nommage, par exemple VP pour Version Provisoire, VF pour Version Finale.

1) DONNER UN NOM BREF ET EXPLICITE (SI VOUS N'ÊTES PAS UNE MACHINE)

MAIS ÇA SE DISCUTE

MES IDENTIFIANTS DOIVENT-ILS ÊTRE OPAQUES OU SIGNIFIANTS ?

On parle d'identifiant **signifiant** quand la chaîne de caractères qui constitue l'identifiant est formée à partir des métadonnées qui décrivent la ressource qu'on identifie.

Au contraire, un identifiant **opaque** est constitué d'une suite de lettres et de chiffres sans rapport avec le contenu de la ressource.

La question du choix entre identifiants signifiants ou opaques est liée à l'utilisation qui est faite de l'identifiant.

Si l'identifiant doit être lu ou manipulé par un utilisateur humain, on aura tendance à vouloir le rendre signifiant.

En revanche, si l'identifiant est principalement manipulé par des machines, la longueur de l'identifiant et son opacité importent peu.

<http://d.opencalais.com/genericHasher-1/c1008719-bca3-3c48-966a-192869743423.html>,

Ou

<http://dbpedia.org/resource/Manhattan>

Il faut être conscient qu'embarquer de la sémantique dans l'identifiant expose à être confronté aux questions d'ambiguïté de la langue. Il sera éventuellement nécessaire de faire porter par l'identifiant des éléments de désambiguïstation pour en garantir l'unicité.

Par exemple, il est nécessaire d'ajouter un élément de désambiguïstation dans l'identifiant pour distinguer l'objet « bidon » (<http://exemple.com/categorie/Bidon>) et la commune « Bidon » (<http://exemple.com/commune/Bidon>).

On voit le type d'ambiguïté que peuvent porter des termes comme « avocat » (homonymie) ou « opéra » (polysémie).

Par ailleurs, les possibles évolutions de la langue induisent un risque pour la pérennité.

Par exemple, le concept qui serait ainsi identifié : « <http://exemple.com/categorie/biodiesel> » devient d'une certaine façon obsolète, dès lors que le journal officiel du 22 juillet 2007 a préconisé l'utilisation de la forme française « biogazole » de préférence à « biodiesel ». Dans ce cas, ou bien l'identifiant du concept s'écarte du libellé du concept, ce qui réduit l'intérêt d'avoir un identifiant signifiant, ou bien on modifie l'identifiant du concept pour l'adapter à l'état de la langue, ce qui remet bien sûr en cause la pérennité de cet identifiant.

Recommandations

Critères	Niveaux d'exigence
Se conformer au plan de nommage de l'institution ou du domaine ou le cas échéant déterminer un plan de nommage spécifiant les éléments de la numérisation à doter d'un identifiant	Obligatoire
Garantir l'unicité des identifiants	Obligatoire
Veiller à la non-ré attribution des identifiants de la ressource concernée	Obligatoire
En cas de suppression ou de dépublication, en renseigner les circonstances et le cas échéant pointer vers une ressource de substitution	Recommandé
Publier sa politique de gestion des identifiants	Recommandé
Donner des identifiants non signifiants pour éviter toute ambiguïté liée aux langues naturelles	Recommandé
Dans le cas de l'utilisation d'identifiants pérennes, prévoir des URI HTTP [c'est-à-dire que les URI donnant directement accès à l'utilisateur, via un navigateur, à une vue de la ressource].	Recommandé

Éléments	Règle	Exemple
Sujet	Obligatoire Il s'agit du sujet principal traité au sein du document. Utiliser des noms communs, écrits en lettres minuscules non accentuées.	projet formation évaluation
Séparateur	Les espaces sont interdits. Utiliser l'underscore (touche 8 du clavier) pour remplacer les espaces	« _ »
Type de document	Facultatif Qualifie la nature du document. Toute abréviation sera en lettres majuscules.	(CR) compte rendu (OJ) ordre du jour
Date	Obligatoire Date de création du document, date de l'événement. Format à l'américaine : AAAAMMJJ. Nommage d'une période : utilisation d'un séparateur « _ » ou « - ».	20180122 201608 2010 201501_07 ou 201501-07
Version du document	Obligatoire Distingue les différentes versions d'un document, signalées par un « V » majuscule suivi de deux chiffres ; version provisoire (VP) et la version finale (VF), version validée (VV). Un nouveau document créé à partir d'une version finale doit être sauvegardé sous un nouveau nom de manière à ne pas écraser la version précédente.	CR_CFVU_V0.0 CR_CFVU_V0.1 CR_CFVU_VP, VF ou VV
Extension	Obligatoire L'extension est ajoutée automatiquement par le système et n'apparaît peut-être pas sur vos écrans.	.txt (fichier texte) .doc (fichier Word) .xls (fichier Excel)