

Date: Lundi 25 Novembre 2019 - 14h-16h

Présents: Aline E., Anaïs O., Esther C., Evelyne S., Grégory M., Katia V., Magali N., Marianne B., Mark A., Patricia V., Pierre F., Soizick L.

1/6 - Nom des fichiers Parfile

- **Actuellement:** AAAA_MMX
où X est le n° de fascicule publié dans le mois...mais *pas le vrai numéro de fascicule dans le volume chez l'éditeur.*
- **Problème:** impossible de générer automatiquement ce n° de manière fiable
- **Proposition:** AAAA_MM_VVVV_FF
 - AAAA: année
 - MM: mois (01-12)
 - VVVV: n° de volume chez l'éditeur
 - FF: n° de fascicule chez l'éditeur
- **Impacts:**
 - plusieurs scripts internes accèdent régulièrement (généralement toutes les nuits) à ces fichiers Parfile => les lister puis les mettre à jour
 - 1 fait à retoucher également
- **Verdict:** OK

2/6 - Séparation ApJ et ApJL

- **Actuellement:** ApJ et leurs Letters sont traités ensemble. C'est à dire qu'ils sont dans le même répertoire et partagent le même parfile.
- **Problème:** chez l'éditeur, il s'agit de 2 journaux différents et ils ne paraissent pas en même temps. Donc il faut toujours exécuter 2 téléchargements, générer 2 Parfile puis les concaténer.
- **Proposition:** les considérer comme 2 journaux différents.
- **Impacts:** *hormis un changement (simplification ?) d'organisation, a priori, aucun*
- **Verdict:**
 - séparation OK...
 - ...mais changement de Bibcode à rediscuter: *faut-il mettre le 'L' devant le numéro de l'article comme maintenant ou utiliser le code journal ApJL ? pour information, avant le passage du numéro de page au numéro d'article le L était dans le 14 caractère du bibcode). Le même problème se posera avec les MNRAS et MNRASL.*

3/6 - Bibcode sans lettre d'auteur dans les XCDS

- **Actuellement:** le Bibcode généré automatiquement avec les auteurs tels que fournis dans le XML.

- **Problème:** en cas d'erreur dans le nom du 1er auteur (e.g. inversion nom de famille et prénom, nom composé, ...), le Bibcode doit être corrigé...partout (dans la hiérarchie, le XCDS et le Bibcode). Si cela n'est fait qu'après mise à jour de Simbad, il faut en plus mettre à jour Simbad.
- **Proposition:** ne pas inclure la 1ère lettre du nom de l'auteur dans le Bibcode utilisé pour les noms de fichiers et répertoire et même dans le XCDS.
- **Impacts:** ambiguïté encore possible avec les IBVS
- **Verdict:** OK, jusqu'à IBVS (*éventuellement voir pour ajouter la lettre R/Q/S... dans le quatorzième caractère*)

4/6 - ATel importés automatiquement dans Simbad

- **Verdict:**
 - OK pour téléchargement automatique des **ATel**, mais cela reste une exception.
 - email de rapport automatique avant mise à jour dans Simbad pour que les auteurs soient vérifiés avant (maj auto après un certain délai...à définir)
 - **Attention:** la vérification actuelle du nom des auteurs est basée principalement sur la liste des auteurs de SIMBAD. Si l'on introduit des erreurs dans SIMBAD, les tests suivants se dégradent et les vérifications ne seront plus valables au bout d'un certain temps !! Donc il est important de vérifier les auteurs avant la maj ou dans la journée qui suit (d'où la nécessité des tests journaliers).

5/6 - Majuscules dans le titre

- **Actuellement:** quelques journaux dont ceux de l'IoP publient des articles où dans les titres chaque mot commence par une majuscule. Pour des soucis d'homogénéité dans Simbad et basés au départ sur les règles de typographie, les scripts de Thomas suppriment les majuscules et les remettent à nouveau en cas de nécessité (noms d'objet, d'instruments, missions, ...). Ces programmes sont basés sur des listes qui se mettent à jour au fur et à mesure et à partir du contenu des abstracts.
- **Problème:** Il reste des erreurs, et les titres doivent être vérifiés systématiquement et corrigés manuellement en cas d'erreur.
- **Proposition:** garder les titres tels que fournis par l'éditeur.
- **Impacts:** *a priori, aucun hormis le fait que l'on perde l'homogénéité*
- **Verdict:** OK

6/6 - Correction des noms d'auteurs

- **Actuellement:** certains noms d'auteurs sont mal rapportés (déjà dans les fichiers XML) : confusion entre le nom de famille et le prénom ou encore les noms à particules ou composés. Les scripts de Thomas incluent un certain nombre de règles pour pouvoir essayer de corriger cela en plus des comparaisons avec ce qui existe dans SIMBAD.
- **Problème:** les corrections sont effectuées manuellement au moment de l'import dans Simbad. Cela prend beaucoup de temps et demande d'effectuer des recherches de l'auteur généralement en se basant entre autre sur la bibliographie de l'auteur présente dans l'article, et éventuellement parmi d'autres sources (e.g. ORCID, ADS) suivant les cas.
- **Propositions:**
 - ne plus corriger les noms d'auteurs (sauf le 1er pour le Bibcode, et les 3 premiers pour les acronymes)
 - garder le système actuel (i.e. exécuter les scripts de Thomas + corrections manuelles systématiques)
 - au moins pour les nouveaux auteurs, effectuer une recherche automatiquement dans ADS afin de donner une proposition de correction

- *(éventuellement complémentaire de la précédente)* vu que ce sont souvent les mêmes noms d'auteur qui doivent être corrigés, les lister quelque part avec toutes leurs écritures possibles *(à mettre à jour régulièrement)* et proposer une correction *(qu'il faudra toujours valider)*
- utiliser ORCID
- **Verdict:**
 - correction en plus de la version originale dans le XCDS (i.e. 2 tags XML: 1 pour l'original, 1 pour la version corrigée par le CDS)
 - donner des suggestions de correction via la page Web d'import ; par ordre d'exécution (i.e. de fiabilité):
 1. Comparaison avec les auteurs déjà présents dans SIMBAD (comme maintenant)
 2. Liste des noms déjà corrigés *(dans de précédents articles ; donc il faut enregistrer automatiquement les écritures possibles de chaque nom d'auteur au fur et à mesure)*
 3. Script de Thomas
 4. ORCID
 5. ADS (mais pas toujours correct non plus)