

Réunion PhyloAlps #3

27/03/2017

Participants

- Eric Coissac
- Frédéric Boyer
- Roland Douzet
- Anthony Hombiat

Modèle de données

- Nomenclature ≠ taxonomie : la nomenclature binomiale contient les deux derniers niveaux de la taxonomie.
- Nomenclature :
 - *Basionyme* : premier nom qui a été donné
 - *Synonyme nomenclatural*
 - *Synonyme de remplacement*
- Concept de `Curator` ≠ `Contributor` :
 - `Curator` : spécialiste missionné par une institution pour renseigner les informations à propos des spécimens d'une collection (par exemple, un herbier) dont il est, par suite, responsable.
 - `Contributor` : Personne autorisée à ajouter des échantillons (`Sample`) dans la base et dont il est donc responsable.
- Le taxid du NCBI sera utilisé comme référence référentiel taxonomique de référence (!) : il sera donc obligatoire au moment de l'ajout d'un échantillon en bd.
- Les id dans d'autres référentiels taxonomiques pourront être renseignés lors de l'ajout d'un échantillon en bd, auquel cas ils seront marqués comme `expertised` . Sinon, et que le système est parvenu à les inférer automatiquement, ils seront marqués comme `inferred` .
- Part d'herbier ≠ échantillon
- Part d'herbier (`HerbPress`) : planche avec un individu de l'espèce qui est enregistré dans un herbier qui fera référence. La part d'herbier est une forme de collection de musée particulière.
- Echantillon (`Sample`) : Tout ou partie d'un ou plusieurs individu(s), pouvant faire (ou non) l'objet d'une part d'herbier.
- Tout ce qui a trait au metabarcoding dans le modèle doit être supprimé : l'objectif du projet PhyloAlps n'est pas d'enregistrer le résultat du metabarcoding mais d'être une base de référence à la disposition des personnes qui le pratique.
- Ajouter la chaîne `SampleSet` > `Sample` > `DNAExtract` > `DNALibrary` > `DNARawSequences` > `DNAAssembly`
- Un extrait d'ADN (`DNAExtract`) ne peut être lié qu'à un échantillon (`Sample`), et pas à une part d'herbier (`HerbPress`) ni à un taxon (`Taxon`).

Prototype

- Plusieurs petites bases de données relationnelles :
 - Taxonomie NCBI sur ncbi.phyloalps.org
 - Taxonomie PlantList sur sajf.phyloalps.org
 - Herbier SAJF sur plantlist.phyloalps.org
 - etc.
- Modèle de données englobant en RDF
- Association modèle relationnel - modèle graphe via une des [implémentations de la R2RML](#)
- Le prototype devra servir de preuve de concept sur une bd de moins de 100 000 échantillons.
- Les mises à jour se feront sous la forme de nouvelles versions délivrées ponctuellement, pas d'intégration continue.

TODO

- Rajouter carte d'échantillonnage PhyloAlps sur phyloalps.org
- Ajouter courriels de contact et courriels des membres du projet via gandi.net
- Faire le lien entre les référentiels taxonomiques de PlantList et du NCBI
- Sur une machine virtuelle NixOS, mettre en place un SGBD PostgreSQL hébergeant la taxonomie du NCBI et l'implémentation R2RML permettant de l'interroger en SPARQL